Measuring Gender Essentialism in Children and Examining the Influence of Indirect Contact with Transgender People on Essentialist Beliefs

by

Rachel Danielle Fine

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2021

Doctoral Committee:

Professor Susan A. Gelman, Co-Chair
Associate Professor Arnold K. Ho, Co-Chair
Professor Kristina R. Olson, Princeton University
Professor Robin M. Queen

Rachel D. Fine

rdfine@umich.edu

ORCID iD: 0000-0001-7784-4891

Dedication

I dedicate this dissertation to my daughter, Olive. Olive, of all my life's accomplishments, being your mom is my proudest and most notable.

Acknowledgements

This dissertation and my graduate career would not be possible without the support of my advisors, mentors, colleagues, undergraduate students, friends, and family. I am grateful to everyone who has helped me with time, advice, and support. I am also grateful for the financial support of the NSF through the Graduate Research Fellowships Program.

First and foremost, I would not have completed this dissertation without the support and guidance from my advisors, Arnold Ho and Susan Gelman. Arnold encouraged me to design rigorous studies and measures and continue to expand my statistical skillset. He pushed me to think deeply and offered advice and guidance on all of my projects. Susan supported me in developing my research ideas and interest in understanding how children think about gender and gender diversity. She also guided me in creating effective and innovative studies to investigate these interests and provided me with many opportunities that will no doubt help me on my career path.

My committee members Robin Queen and Kristina Olson have been wonderful to have on my committee and equally wonderful to have as collaborators on timely research that is much needed. I am so grateful for your advice, feedback, and time over the past couple years. Robin's perspective as a linguist has added richness to this dissertation and our collaboration on pronoun usage. Kristina has given me very useful career and funding advice in addition to our collaborations.

Kristina, Susan, and Selin Gülgöz included me in the NIH grant funded project that allowed me the opportunity to collect data for this dissertation testing the reliability and validity of the GES-C and that I continue to enjoy collaborating on. Kristina and Selin also gave great feedback to help develop the items of the GES-C.

I am so thankful to have worked with a wonderful group of undergraduate students and lab managers, many of whom helped collect data for this dissertation. I hope they were able to learn from me because I learned so much from them. Logan Burley, Apoorva Asthana, Kathryn Chang, Madelyn Brunvand (recruiter extraordinaire), Athena Bryer, Kaleah Brauher, Keegan Giffels, Sam Heiman, Shannon Jajko, Vanessa Lazaro, Sally Kafelghazal, Lexi Nylander, Sally Valcarcel, Lyndsey Van, Zu Weinger, Maria Camila Samboni, Colleen Tacubao (the wonderful artist of Jazz and Blue) and Aly Gonzales (the voice of Jazz and Blue) you were invaluable to my experience at UM. Former lab managers who went on to great things -Nicole Cuneo, Valerie Umscheid, and Hannah Meloche- were instrumental in collecting the data for this dissertation and I enjoyed all the time we shared driving all over rural Michigan.

I deeply value all that I learned in lab meetings with Henry Wellman, Felix Warnken, Ioulia Kovelman, and Alexandra Rosati. They offered advice that improved my research and presentations. I also appreciated all the feedback from past and present graduate students and postdocs including, Danielle Labotka, Giulia Elli, Kristan Marchak, Young-eun Lee, and Sarah Probst.

I also received overwhelming support from my cohort and fellow graduate students. Zach Reese, Yuyan Han, Cristina Salvador, Qinggang Yu, and Ramona Perry are individually, and as group, the best people one could hope to go through a graduate program with. Nicholas Michalak, Koji Takahashi, Megan Seymour, and Leigh Goetschius are supportive friends that

helped make my time in Ann Arbor so enjoyable. Also, Zach, Nick, and Koji along with our significant others, are a trivia force to be reckoned as long as there are no sports questions.

Additionally, I formed meaningful friendships and received encouragement from Alana Pawlak, Paul Irving, Mary Irving, and Paul Hamerski. You are not only some of the coolest physicists I've ever met, but some of the coolest people full stop.

My family has always been loving and supportive part of my life. My parents, Bill and Sarah, and my brother, Nolan, have always supported me in my academic endeavors and have greatly shaped the person I am today. My grandparents -Nancy, Bob, Shirley, and Dan- always made me feel deeply loved as did my aunts, uncles, and cousins.

And the person I most want to thank most is my husband, Abhilash. Life with Abhilash and Olive is full of love and happiness. Abhilash celebrated my achievements and comforted me through disappointments. He even helped with coding and stats questions. These past five and a half years have been one of the best periods in my life and that is in large part due to Abhilash.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Recently, gender diversity has become more visible in the U.S. Yet many still struggle to understand gender identities outside of the binary of man and woman (Buck, 2016). One lay theory children and adults may use to think about gender and specific gender identities is essentialism. Essentialism is a set of beliefs that center around the idea that certain categories have an unknown or ill-defined essence. As a result of this presumed essence, the categories are thought to be biologically based, discrete from one another, informative about category members' behaviors and preferences, and immutable. Although prior research has established the use of essentialist beliefs about gender from an early age, several questions remain, especially at a time when gender diversity is becoming more visible. In this dissertation, I: (1) developed a new scale of gender essentialism for children five to ten years of age, the Gender Essentialism Scale for Children (GES-C); and (2) examined the effect of stories about trans-identity characters on children's understanding of transgender identities and gender essentialism.

The GES-C is a 16-item measure of gender essentialism with four four-item subscales measuring the components of essentialism described above. I found the GES-C to be a reliable and valid scale with 316 participants aged five to ten years old. I also performed a confirmatory factory analysis (CFA) using structural equation modeling (SEM) and found my scale to have fit indices outside of commonly used cutoffs for good model fit but in line with the other scales for children specifically developed for use in developmental research psychology.

Next, I conducted a study with 173 five- to six-year-old and nine- to ten-year-old children to test what children can learn about transgender identities from stories and whether this can lead to a reduction in gender essentialist beliefs. Participants in this study were assigned to one of three conditions, varying in the story that they heard: the *Realistic story* about a transgender girl socially transitioning from a boy to a girl; the *Metaphorical/Fantastical story* about an anthropomorphized, red-labeled marker who discovers their identity as blue (this story could be interpreted as a metaphor for being transgender); or *No story* (control). Hearing the realistic story about the transgender girl significantly improved understanding of transgender identities. And although I found no overall reduction in gender essentialism, essentialist beliefs about the immutability of gender were reduced after hearing the realistic story.

These findings underscore the importance of examining gender essentialism, wholly and by component, in children. Being able to efficiently and effectively measure multiple components of gender essentialism at one time allows researchers to better measure when and how essentialist beliefs change in children. It will be especially important to understand how children's gender essentialist beliefs may or may not change as a result of the increased visibility of gender diverse identities.

Chapter 1 Introduction

How we think about gender is ever evolving. Trans, nonbinary, and gender nonconforming people are beginning to be recognized in entertainment, politics, and everyday life. Lilly and Lana Wachowski, Laverne Cox, Elliot Page, and Jazz Jennings are all trans celebrities that have become household names. The state of Delaware elected the first transwoman, Sarah McBride, to a U.S. state senate, in 2020 (Flores et al., 2020). In 2017, Canada began allowing citizens to identify outside the gender binary of man and woman on their passports (Chokshiaug, 2017). And, in the field of psychology, researchers are beginning to be inclusive of all genders in their work, differing from the past when researchers in the field largely treated gender as a binary category interchangeable with sex (Hyde et al., 2018; Martin, 1989; van Anders, 2015). With all these changes, researchers are presented with a unique opportunity to examine how children conceptualize gender at a time when past notions of this social category are being challenged. Focusing on children's conceptualization of gender is important given the salience of gender in childhood (Arthur et al., 2008; Bem, 1983). Children view gender as significant earlier than other social categories such as race (Davoodi et al., 2020; Shutts et al., 2013). And across many cultures, children *essentialize* gender (Davoodi et al., 2020; Diesendruck et al., 2013; Diesendruck & Weiss, 2015)—that is, treating gender differences as natural, objective, unchanging, and predictive of a host of properties. In the years to come, it will be fascinating to see if children continue to essentialize gender in the same ways as they have done in the past. However, a barrier exists to fully measuring these changes, as currently there is

no single measure that comprehensively examines essentialist beliefs in children. To address this gap in the literature, I developed a new measure of children's gender essentialism, the Gender Essentialism Scale for Children (GES-C), and use it to examine how acknowledging gender diversity may influence children's beliefs about gender.

Essentialism, a broad lay theory used to varying extents by adults and children, is a way to categorize the natural world. I emphasize that this is a lay theory, because essentialist beliefs are often shown to be scientifically incorrect in relation to social categories (for an example of scientific evidence against gender essentialist beliefs see Ainsworth, 2015). Essentialism is an intuitive belief about certain categories, that members of the category have an underlying reality that is inborn and unchanging. For example, in the case of gender, essentialism is the belief that boys and girls are inherently, biologically distinct, that there is a sharp boundary between boys and girls, and that one cannot change from one gender to the other. Essentialism also involves a belief that there is an inherent "essence" to boys or to girls that makes them what they are and influences their behaviors and preferences. Importantly, however, what that essence is might not be known (for example, children might believe that there is *something* biological, internal, and non-obvious that differentiates girls and boys, but not have any specific idea of what that is). In that sense, essentialism can be a "placeholder" concept that may or may not become more well-defined for a person (Gelman, 2003; Medin, 1989; D. Medin & Ortony, 1989). A variety of studies indicate that children essentialize the social category of gender (Martin et al., 2002; Rhodes & Gelman, 2009a; Taylor, 1996; Taylor et al., 2009).

Importantly, it is generally recognized that essentialism is not a single construct but rather can be conceptualized as having multiple strands or components (Bastian & Haslam, 2006; Gelman, 2003; Gelman et al., 2007; Rhodes & Mandalaywala, 2017). Importantly, these

components have been shown to be empirically distinct, especially in young children, and to have distinct consequences (Gelman et al., 2007; Rhodes & Mandalaywala, 2017). Although essentialist beliefs have been labeled differently and parsed into components with varying levels of specificity, research has consistently demonstrated the following four components: biological basis (i.e., categories are natural kinds and reflect inborn nature), discreteness (i.e., categories have sharp or strict boundaries), informativeness (i.e., categories are homogeneous and have inductive potential), and immutability (i.e., categories are stable over time and contexts; Gelman, 2003; Rhodes & Mandalaywala 2017). These four components will be the focus of my Gender Essentialism Scale for Children (GES-C) and, subsequently, the focus of the research conducted for my dissertation. I briefly describe each of the components below, and how they have been examined in prior research.

Biological Basis

One component of essentialism is the belief that a category has a biological basis. This does not necessitate a precise set of beliefs about biological mechanisms (such as genetics or DNA), but rather an assumption that internal, bodily features are more important than outward features in determining category membership, and that category members have an innate disposition to develop category-typical features. By four years of age, children endorse the idea that internal physical features of an animal or person are important to its category membership (e.g., a dog without its insides is no longer a dog, whereas a dog with its fur shaved off is still a dog; Diesendruck & Weiss, 2015; Gelman & Wellman, 1991). Four-year-olds also report a sense of innate potential, in that how a newborn animal or plant seed grows is determined by its origins rather than its environment. For example, a baby rabbit raised by monkeys will grow up to like carrots, not bananas, or an apple seed planted in an orange orchard will grow into an apple tree,

not an orange tree (Gelman & Wellman, 1991). With respect to gender, children demonstrate a strong belief that gender is biological by affirming that the gender assigned to a child at birth will coincide with that child's biological sex when they are older. In one study, children endorsed statements that an infant's gender assigned at birth would match the "gender" of their blood, body, brain, and heart at age ten (e.g., a girl infant would have girl blood at age ten) at levels well above chance (Taylor et al., 2009).

Discreteness

A second component of essentialism is the belief that categories are discrete—that is, that there is a sharp boundary between different categories, and these boundaries are assumed to be objective. This is not to say that an individual, animal, or plant cannot belong to multiple categories, but that boundaries are "intensified", or deemed relatively more important (Gelman, 2003). Illustrating the importance of boundaries for natural kinds (categories assumed to be found in nature, such as animal species or distinct types of substances) as opposed to artifacts (such as furniture or vehicles), Diesendruck and Gelman (1999) found that adults judged animals as more absolutely a member of their category than artifacts were of their category, regardless of item typicality. For example, adults judged that a penguin and a sparrow are each fully birds, whereas a clock is less of a piece of furniture than a desk. Even children as young as three believe that there are boundaries distinguishing between categories of animals regardless of the typicality of the animal being categorized (Rhodes et al., 2014; Rhodes & Gelman, 2009b). Furthermore, children treat boundaries as objective, such that members of different categories (e.g., dog vs. cat; girl vs. boy) are distinct and not open to debate (Rhodes & Gelman, 2009a; Rhodes, Karuza, & Gelman, 2014). In fact, young children view gender as particularly discrete when compared to other social categories. Children as young as three years of age were as likely

to treat a girl and a boy as different kinds, as they were to treat a cat and a dog as different kinds. These same children did not draw as sharp boundaries between children of different races (Rhodes et al., 2014).

Informativeness

A third component of essentialism is the belief that a category is informative. In other words, knowing the category that an individual belongs to licenses a rich array of inferences about altogether new properties (also known as inductive potential; Gelman, 2003). Children as young as two years of age make novel inferences about natural kinds based on category membership more than perceptual similarity (Gelman & Coley, 1990; Gelman & Markman, 1986, 1987). And, children, regardless of their gender identity of cisgender (gender identity of a person who identifies as the sex and gender assigned to them at birth) or transgender, use sex at birth to make predictions about a child's preferences and behaviors, even when sex at birth is pitted against environmental rearing conditions (Gülgöz et al., 2019; Taylor, 1996; Taylor et al., 2009). This task also illustrates how different components may be interrelated and combined in one measure, as it simultaneously assesses informativeness as well as the biological basis of gender.

Immutability

Finally, a fourth component of essentialism is the belief that a category is immutable. For example, by 3 years of age, children understand that wearing a costume does not change a lion into a tiger, and by 7 years of age, children understand that a lion cannot be transformed into a tiger by a scientist who performs an operation to change how it looks (Keil, 1989a). By about 6 or 7 years of age, children endorse similar beliefs about the stability of gender/sex category membership (Bem, 1989; Gouze & Nadelman, 1980; Szkrybalo & Ruble, 1999), although this

belief increases over early childhood, in part due to increasing exposure to religious or scientific beliefs that support immutability. For example, children who assert that gender is not determined by clothing, hairstyles, or behavioral preferences (e.g., liking to cook) are more likely to assert that gender is immutable by using a rationale grounded in biology (e.g., “boys have penises”) or religious beliefs (e.g., "God made you a girl"; Szkrybalo & Ruble, 1999). Nonetheless, even younger children, by 3 years of age, report that the sex of a baby does not change when the baby is dressed in a way that does not match their sex, if anatomically correct information is provided (Bem, 1989).

Development of Gender Essentialism in Childhood

Children’s gender essentialism follows a developmental trajectory that has been well-established in the literature. By preschool age, children express an inflexibility in their thinking about gender and rely heavily on essentialist beliefs and stereotypes (Arthur et al., 2008; Bem, 1983; Bigler & Liben, 2006; Martin et al., 2002; Taylor, 1996; Trautner et al., 2005). Numerous studies have found that children’s beliefs about gender become less rigid as they age and that children show a marked reduction in essentialist beliefs about gender around nine or ten years old (Davoodi et al., 2020; Taylor et al., 2009). At the same time, these developmental patterns differ by social context. In a study of two communities in Michigan just 75 miles apart from one another, Rhodes and Gelman (2009) found that children in a conservative, rural town were more likely than children from a liberal, university city to treat gender boundaries as objective. Moreover, on this task, children’s gender essentialism lessened with age in the university town, whereas it remained stable from kindergarten through twelfth grade in the rural community. These differences suggest that by elementary school age, children are picking up on environmental cues regarding gender. Given this early sensitivity, an opportunity arises to teach

children about some of the more limiting aspects of gender essentialist beliefs, specifically regarding gender diversity.

Gender diversity includes gender identities that do not strictly adhere to binary gender (man or woman) norms and roles assigned to an individual at birth. These gender nonconforming identities can encompass transgender, genderfluid, and nonbinary identities, among others. Some of these identities encompass all or many genders within one identity; others involve a change from the binary gender assigned to a person at birth to the other binary gender. These types of nonbinary identities introduce nuance into the concept of gender. This moment in time, when these diverse identities are becoming more visible, gives psychologists an opportunity to better understand if and how ideas about gender, and possibly social categories more generally, can shift from a binary construct almost inseparable from biological sex, to a more varied construct further removed from sex.

The Present Studies

The issues I address in my dissertation research are how we can more accurately and comprehensively measure children's gender essentialism, and, with this new means of measuring essentialism, how learning about transgender identities relates to children's gender essentialist beliefs. In examining these questions, I construe gender essentialism in terms of the four components outlined in the prior section. Over the course of two studies, I created a new gender essentialism scale for children aged five to ten years old, and tested if exposure to transgender identities through direct and metaphorical stories reduces, increases, or has no effect on gender essentialism in five- to six-year-old and nine- to ten-year-old children. Below I briefly sketch out the motivation and research questions for each of the dissertation chapters.

Chapter 2: Developing the Gender Essentialism Scale for Children (GES-C)

The goal of Chapter 2 was to develop a scale that can measure children's essentialism grounded in the four components of essentialism previously discussed. My goal was to provide researchers a scale with the flexibility to measure essentialism as a whole when using the scale in its entirety, as well as individual components with sub-scales. I also wanted to better capture the nuance with which children view gender. Prior research does not provide a measure for children that is both developmentally appropriate and broad enough to examine multiple components at once. In past research, measures were often created for a specific study to look at just one or two components of essentialism, and thus could not examine gender essentialism broadly. Comprehensive scales have been used with adults (e.g., Haslam et al., 2000; Lee et al., 2020), but these scales are too complex for use with young children. My measure includes four sub-scales that can individually measure biological-basis, discreteness, informativeness, and immutability. This new scale may give researchers the ability to better compare gender essentialist beliefs of children and adults and their relation to other relevant psychological concepts.

Chapter 3: Indirect contact with transgender identities through stories and its influences on children's gender essentialism

The goal of Chapter 3 was to examine the effect of stories about transgender identities on children's gender essentialism, making use of the scale developed in Study 1. Study 2 included two different story conditions: one focused on the social transition of a binary transgender girl (i.e., a child who was assigned a boy at birth who socially transitions to a girl), and the other was a metaphorical version of the same story. In the second story, the character is an anthropomorphized marker with blue ink on the inside and a red plastic casing on the outside, who likewise goes through a social transition (from a marker that was assigned red at first who

transitions to a blue social identity). In a third (control) condition, child participants did not see a story. In all three cases, I then measured children's essentialist beliefs about gender. Children are relatively unlikely to come into contact with many transgender people in everyday life, given that trans and gender-nonconforming people are currently estimated to account for 0.39% of the population (Meerwijk & Sevelius, 2017). However, through stories, children can "meet" a transgender character. I predicted that, after learning about a transgender character, children would endorse gender essentialist beliefs to a lesser extent because transgender identities challenge the beliefs related to the four components I examined with the GES-C. I also predicted that the realistic story would be more effective than the metaphorical story, given children's difficulties in extending metaphors as intended (Richert & Smith, 2011).

Chapter 2 Developing the Gender Essentialism Scale for Children (GES-C)

As I highlighted in the previous chapter, there is a long, rich history of studying gender essentialism in children (Gelman, 2003; Rhodes & Mandalaywala, 2017). However, extant measures of children's gender essentialism often include only one or two components of essentialism. This makes it cumbersome and time consuming to measure multiple components of essentialism within a single study. And, because there is evidence that the components of essentialism are distinct but interrelated (Rhodes & Mandalaywala, 2017), a comprehensive measure would be useful in further exploring the development of each component individually and in relation to one another. With my new scale, the Gender Essentialism Scale for Children (GES-C), I offer a more nuanced and comprehensive way to measure gender essentialism in children as young as five years old.

Existing Essentialism Measures

Because essentialism is a complex system of beliefs, researchers often focus on one or two elements of essentialism in a study. I will discuss how each component of my new scale has been measured in the past, as well as the value of an additional, more comprehensive measure that can be conducted in one brief session.

Biological Basis

Measuring the belief that categories have a biological basis presents researchers with the challenge of asking young children about biology in terms they understand. Whereas adults have knowledge about DNA and other unseen but important biological concepts, children often have

at most minimal exposure to these concepts. Gelman and Wellman (1991) addressed this issue by examining children's beliefs that for certain categories, *insides* are more important than *outsides*. One of their studies demonstrated that children as young as four years old privilege the unobservable insides of essentialized categories and other insides-relevant objects to outward appearances by asking children if certain things were still the same type of things if their insides were removed or their outsides were removed. For example, children were asked about the insides removal of a dog, "What if you take out the stuff inside of the dog, you know, the blood and bones and things like that and got rid of it and all you have left are the outsides?" and the outsides removal of a dog, "What if you take off the stuff outside of the dog, you know, the fur and got rid of it and all you have left are the insides?" After each type of removal children were asked, "Is it still a dog?" and "Can it still bark and eat dog food?" Children typically indicated that the insides were more critical than the outsides (e.g., it would no longer still be a dog and bark if it lost its insides, but it would still retain these features it lost its outsides). This work provided evidence that children endorsed an essentialist framework about animals and were not limited to merely evaluating outward appearances.

Discreteness

Children who endorse discreteness beliefs about categories believe that categories have sharp boundaries, and/or that category boundaries mark distinctly different kinds. They believe a penguin is as much of a bird as a robin even though penguins cannot fly. This was demonstrated by the measure that Rhodes and Gelman (2009b) developed. In their measure, children were presented with typical (e.g., robin) and atypical (e.g., penguin) members of a category (e.g., birds) and a nonmember (e.g., butterfly) that shared a trait with typical members of the category. Participants were asked if each item (e.g., typical, atypical, and nonmember examples) was a

member of the category (e.g., “Is a robin a bird?”). Participants could answer thumbs up/definitely, thumbs to the side/sort of, or thumbs down/definitely not a category member. Children were more absolute about categories of natural kinds (e.g., birds) than artifacts (e.g., tools).

Another task of discreteness measured children’s assessment of alternative categorizations. In this measure, children 5-18 years of age were introduced to a visitor from a place far away, named Feppy, who was unfamiliar with “the way we do things” (Rhodes & Gelman, 2009a). In one variation of this measure, the child saw two pictures at a time (e.g., a girl and a boy) and learned that Feppy and all his friends said that they were the “same kind.” Children were then asked, “Are they maybe right?” In this task, when the two pictures were of different categories, a response that Feppy and all his friends were wrong was interpreted as endorsing the belief that discrete category boundaries were objective. This measure was later adapted to examine category boundaries of children as young as 3 years old (Rhodes et al., 2014). These tasks demonstrated that children drew sharper boundaries between genders (boy/girl) and between animals (e.g., cat/dog) than they did for races (e.g., white/Asian) and artifacts (e.g., table/bookcase).

Informativeness

To capture informativeness, Gelman and Markman (1986) showed four-year-old participants sets of three items, all natural kinds (e.g., two squirrels and one rabbit). Participants were told a different behavior or trait for two of the pictures (e.g., this squirrel eats bugs, this rabbit eats grass). Then, participants were shown an item that belonged to the same category of one of the pictures but resembled the other picture (e.g., a kaibab squirrel with rabbit-like ears, referred to as a squirrel) and asked what behavior or trait the thing in the third picture would do

or have (e.g., eat bugs or grass). Children relied on category labels to draw inferences about the third picture at levels above chance, demonstrating that children relied more heavily on category information than outward appearance (see also Booth, 2014; Davidson & Gelman, 1990; Gelman & Coley, 1990; Gelman & Davidson, 2013; Graham & Diesendruck, 2010; Jaswal et al., 2009; Tarlowski, 2018). Outward appearance gained importance for children only when children perceived a causal link between this appearance and an internal property (Diesendruck & Eldror, 2011). This paradigm has been used to look at the inductive potential of gender, ethnicity, and nationality (Diesendruck & HaLevi, 2006; Gelman et al., 1986; Hussak & Cimpian, 2019).

Immutability

Immutability refers to the stability of a category over time. Keil (1989) measured this component with children in kindergarten, second, and fourth grades. Children were told about natural kinds and artifacts that underwent an inward and outward transformation so that they appeared more similar to another animal, artifact, or mineral. Animals had fur changed and surgery to transform them (e.g., raccoon was changed to resemble a skunk via dyed fur and sac of odor inserted inside) and artifacts had parts removed, added, and changed to resemble and potentially function as another artifact (e.g., a coffeepot was changed to resemble a birdfeeder, as its handle was removed, a window was cut into it, the top was sealed, and a pole and bottom plate were added). Children were asked what natural kind or artifact the target was after the transformation and were prompted to explain their choice. Increasingly with age, children asserted that natural kinds retained their original identity after the primarily superficial transformation (e.g., second-graders but not kindergarteners were more likely to say the raccoon was still a raccoon after the surgery). Children across all grades asserted that artifacts changed their identities after transformation (e.g., the coffeepot became a birdfeeder).

Another way immutability has been captured for categories such as gender and race is to present a child participant with pictures of a child and two adults (Kinzler & Dautel, 2012; Pauker et al., 2010, 2016; Roberts & Gelman, 2016; Ruble et al., 2007). One adult matches the child on one dimension (race, gender) and the other adult matches the child on another (e.g., language, emotion). Participants are asked which adult the child will grow up to be. Participants see several triads of pictures and the more the participant chooses the adult that matches the child in the category of interest, the more the child endorses immutability beliefs about the category. Results from these studies showed that generally children believed gender does not change over time at the highest levels by age 5 but race immutability takes more time to develop, specifically for White children in the U.S.

Slaby and Frey (1975) developed a 14-item scale to measure gender constancy, including three components: gender identity (a child can apply the labels 'boy' and 'girl' to themselves and gender-typical others), gender stability (a child's gender is the same at different points in time; e.g., as an adult, as an infant), and gender consistency (a child does not change gender even if they exhibit gendered preferences and wears gendered clothing different from their gender) . An example of the three gender consistency items is, "If you played (binary gender different from participant's binary gender) games, would you then really be a boy or really be a girl?" An example of the two gender stability questions includes, "When you grow up, will you be a mommy or a daddy? Could you ever be a (parent identity child did not give to the prior question)?" And, an example of the nine gender identity questions is, "Are you a girl or a boy? Are you a (binary gender different from participant's response to prior question)?" Slaby and Frey (1975) found that gender constancy developed in stages, with the gender identity component developing first, followed by gender stability, and, finally, gender consistency.

Variations of this scale have been widely used (Bem, 1989; Emmerich et al., 1977; Gouze & Nadelman, 1980; Levy & Carter, 1989; D. N. Ruble et al., 1981; Ruble et al., 2007; Szkrybalo & Ruble, 1999) and with children as young as 27 months old (Levy & Carter, 1989). Research shows children achieve the highest levels of gender stability around 5 years old and reach highest levels of gender consistency between ages 5 and 7 years old (Ruble et al., 2007). This difference may be a result of the elements of informativeness present in the consistency questions (e.g., signaling gender through clothing and games). This scale includes multiple possible signs that gender can be perceived as changing such as behaviorally (incorporating elements of informativeness) and physically, and mixes evaluations of immutability of oneself and specific others.

To differentiate immutability from informativeness in the GES-C, this dimension focuses on change over time (gender stability). Also, because gender identities today include those that would be classified as gender "inconsistent" (e.g., genderfluid identities), I chose to avoid items that may prime a participant to use a binary gender construal such as, "If you played girl games, would you really be a boy or a girl?" because it limits their responses to boy or girl but not both or neither.

Combined Scales

I am aware of two measures with children that address multiple components of essentialism within one scale, though importantly, neither measure separates out the different components. One measures is the Island Task (Taylor, 1996; Taylor et al., 2009). In this task, children are introduced to a hypothetical child who was sent to live on an island with a relative. In one case, an infant girl was sent to live with her uncle on an island where there are only boys and men. In the other case, an infant boy was sent to live with his aunt on an island where there

are only girls and women. Participants are then told that the child is now 10 years of age and asked about the toys the child would prefer (e.g., dolls or tools and nails), activities they would prefer (e.g., put on make-up or go fishing), careers they would prefer (e.g., ballerina or football player), roles they can have (e.g., mommy or daddy), and physical features they would have (e.g., girl's brain or boy's brain). These questions reflect the components of informativeness and biological basis. The more often the participant chooses answers typical for the child's assigned gender, the more they endorsed essentialist beliefs.

Diesendruck and Haber (2009) created a scale in which they compared exemplars of animal, gender, race, ethnicity, socioeconomic status, and artifact categories. An example of the animal categories would be elephants and lions. They then asked questions such as "To what extent do elephants and lions differ in…", where the attribute included: 1) what they like, 2) what they think, 3) how they behave, 4) how they look, 5) what they have inside their body. They also asked two questions about heredity transformations, for example: "Is it possible that a couple of lions will give birth to a baby elephant?" and "If a baby is born to, and raised by, a lion family, is it possible that it will be an elephant when it grows up?" These questions cover the components of informativeness, biological basis, and immutability. First and fifth graders from orthodox and secular schools in Israel answered these questions on a four-point scale, with higher scores indicating more essentialism. Consistent with past literature, both orthodox and secular Jewish children did not endorse all components to the same extent within a category; however, they did essentialize animals and did not essentialize artifacts or socioeconomic status. Although both the Island Task and Diesendruck and Haber's scales measure essentialist beliefs more broadly, neither were created to be able to evaluate individual components of essentialism.

The Gender Essentialism Scale for Children (GES-C)

I endeavored to create a measure comprised of four components of gender essentialism: biological basis, discreteness, informativeness, and immutability. I based my items on essentialism scales used with adults (Bastian & Haslam, 2006; S. R. Levy et al., 1998), but reworded to be understandable to children as young as five years of age. For example, adults respond to the biological basis item, "The gender that people exhibit can be traced back to their biology." My child scale has a similar item that discusses biology in more child-friendly language acknowledging visible signs of internal biological features, "You know a kid is a girl or boy because of what their bodies look like." Although this question references outward physical features, they are outward physical features that children strongly associate with the biology of sex (e.g., penis, vulva and vagina; Bem, 1989). A discreteness item for adults includes, "A person either is a certain gender or they are not," and an analogous item for children is, "Girls and boys are opposites." Like the adult scale, my final scale includes at least one reverse-coded item in each component.

I aimed for a measure that can be used in its 16-item entirety or by each 4-item component. After piloting several items (see Appendix A), I selected 16 items to comprise the final scale. With the GES-C, I am able to measure gender essentialism on a continuous scale (scores on the GES-C range from 1 to 4, with 4 representing the most gender essentialist score) and with consistency (all components are measured in the same way), to build on the important literature on children's understanding and conceptualization of gender.

Method

Participants

The 316 participants were children aged five to ten years old (M = 8.33 years, SD = 1.58 years). One hundred seventy-one were girls (54.1%) and 145 (45.9%) were boys. No parent reported having a child with another gender. The majority of participants were White (N = 229, 72.5%), 33 were multiracial (10.4%), 5 were Asian (1.6%), 3 were Black (0.9%), and 2 were Native American (0.6%). I do not have race information for 44 (13.9%) of participants.

Participant responses were compiled from three different studies. Study 1a was conducted at children's museum and in a psychology lab at the University of Michigan (N = 12). Study 1b was conducted in lab at the University of Michigan and online via a secure University of Michigan Zoom account (N = 52). Study 1c was conducted in a lab at the University of Washington, in a lab at the University of Michigan, in rural, conservative towns in the state of Michigan, and online via Zoom (N = 252).[1] There were no differences in gender essentialism scores between the three study samples (ps ≥ .19).

Materials

Below are the items that were used in the final version of the scale. Please see Appendix A for the instructions and training questions given to participants.

This scale references the genders of girl and boy in its items. In some items both genders are referenced (e.g., girls and boys) and, in some items, only one gender is referenced. As a result, there are 4 versions of the scale, varying both the gender used in single-gender items and the order of genders in the two-gender items:

1) "Girls" for single-gender items; "girls and boys" for two-gender items
2) "Boys" for single-gender items; "girls and boys" for two-gender items

[1] Differences in gender essentialism between urban and rural samples of participants are being examined in another project.

3) “Girls” for single-gender items; “boys and girls” for two-gender items

4) “Boys” for single gender items; “boys and girls” for two-gender items

Biological Basis

Bio1 Girls do girl things because they have girl bodies.

Bio2 You know a kid is a girl or boy because of what their body looks like.

Bio3 A girl is a girl, even before she is born.

Bio4 A kid can choose to be a boy or a girl; it doesn’t matter what their body looks like (reverse-coded).

Discreteness

Dis1 A kid can feel like a girl sometimes and a boy other times (reverse-coded).

Dis2 A kid can feel like a girl and a boy at the same time (reverse-coded).

Dis3 Girls and boys are the same (reverse-coded).

Dis4 Girls and boys are opposites.

Informativeness

Info1 Girls all over the world act the same.

Info2 It's easy to tell if a kid is a girl or a boy by how they act.

Info3 It's easy to tell if a kid is a girl or a boy by the toys they like.

Info4 Girls all act very different from one another (reverse-coded).

Immutability

Imut1 A kid can change from a girl to a boy (reverse-coded).

Imut2 A girl can become a boy if she wants (reverse-coded).

Imut3 If a kid is a girl when they're born, they will be a girl when they grow up, too.
Imut4 A boy can't change that he is a boy.

Measures

Gender Essentialism Scale for Children. The final version of the scale is 16 items with 4 items per component (see Materials). Participants were first asked if they agreed or disagreed with each statement and then whether they agreed/disagreed a little or lot. The responses to these two questions resulted in a scale ranging from 1 “disagree a lot” to 4 “agree a lot.” Final scores for the overall the GES-C and each component were calculated by averaging responses to items, after appropriately accounting for reverse-coded items. Final scores ranged from 1 to 4, with 4 representing the most gender essentialist score on the GES-C.

Island Task (Taylor et al., 2009)**.** In this task, participants were told about a baby girl who moved to an island with their uncle, or a baby boy who moved to an island with their aunt. The island is populated by adults and children of the binary gender different from the baby (e.g., the baby girl lived on an island with only boys and men). Participants were told the baby never meets anyone of their gender (“girl” or “boy”). Participants then answered three questions to check that they understood who lives on the island. Next, participants were told the child is now 10 years old and were asked questions about the toy preferences (e.g., toy truck or tea set), activities (e.g., fishing or putting on makeup), and future career aspirations (e.g., preschool teacher or construction worker) of the child. They also answered questions about biological traits the child would have (e.g., boy brain or girl brain) and the gender-based roles they could have (e.g., husband or wife). There were 15 items in total. Responses that matched the gender assigned to the target character (were stereotypical or expected based on the gender of the target

character) were scored as "1", and responses that matched the child's environment (were stereotypical or expected based on the gender of the other people on the island) were scored "0." Scores could range from 0 to 15, with higher numbers indicating more essentialism.

Procedure

Because participants were recruited from three different studies, I will report the procedure from each study. All participants completed the scale with an experimenter who read the instruction, training items, and scale items to the participant. In the first study (Study 1a) I recruited participants from a local children's museum and a lab database at the University of Michigan, and children completed a version of the gender essentialism scale and then completed the Island Task. In the second study (Study 1b), I recruited only participants who were in the control condition of the experiment reported in Study 2 of this dissertation. These participants answered two questions about whether a child could have the body of one binary gender but feel like the other binary gender. Afterward, participants completed a version of the gender essentialism measure and the Island Task. These participants completed the study in the same University lab as Study 1a, or via an online meeting on Zoom. In the third study (Study 1c), participants completed the GES-C and two other measures related to gender nonconformity in a random order. These participants completed the study in-person at the University lab, at their homes, or in a public library in their hometowns, or remotely on Zoom. In all studies, participants were randomly assigned to a version of the gender essentialism scale.

Results

I will present multiple analyses to assess the reliability and validity of the GES-C. To evaluate if it is acceptable to analyze all participants as one sample, I first examined if there were

differences in gender essentialism scores by study. I also conducted analyses to see if there were differences in gender essentialism scores by version that may be useful to know when using the GES-C in future studies. I was also interested in any potential differences by participant gender or age (see Smiler & Gelman, 2008, for participant gender differences in gender essentialism among adults). Past research suggests boys may endorse gender essentialism more than girls. It is also well documented in the literature that gender essentialism decreases with age in children (Davoodi et al., 2020; Taylor et al., 2009). This was followed by the confirmatory factor analyses (CFAs) with and without modifications to number of items used and errors correlated to assess how well the GES-C measures the four components as separate factors. Next, I presented the internal reliability of the scale (Cronbach's alpha). Finally, I analyzed the correlations of the GES-C with the Island Task to address the validity of the scale in comparison with existing measures of gender essentialism.

Comparisons of the GES-C Scores by Study Sample and Scale Version

As noted previously, scores on the 16-item full scale did not significantly differ by study ($ps \geq .19$). They did not vary by study for the biological basis component ($ps \geq .79$), the informativeness component ($ps \geq .08$), or the immutability component ($ps \geq .14$). They did vary on the discreteness component. Scores in Study 1b and Study 1c were significantly lower than scores in Study 1a (Study 1b vs. Study 1a, $t(62) = -2.90$, $p = .01$; Study 1c vs. Study 1a, $t(261) = -2.53$, $p = .01$)[2] but were not different from one another ($t(301) = -1.18$, $p = .24$, see Table 2-1 for means and standard deviations). Given that these samples varied on only one component, and

[2] These differences were not due to age, as participants' age did not significantly vary by sample. These differences were also not due to outliers in Study 1a, as there were no participants in that study whose scores were 2 standard deviations above or below the mean of discreteness across all studies ($M = 2.35$, $SD = .63$).

that the two larger samples were consistent with one another, I collapsed across all three datasets when conducting analyses.

Table 2.1 Means and Standard Deviations for the GES-C and its Subscales by Study

	Study 1a	Study 1b	Study 1c
16- item Full Scale	2.52(.48)	2.28(.57)	2.34(.57)
Biological Basis	2.44(.75)	2.46(.77)	2.49(.75)
Discreteness	2.81(.60)	2.24(.62)	2.35(.62)
4-item Informativeness	1.69(.47)	1.79(.62)	1.84(.61)
Immutability	3.13(1.07)	2.63(1.08)	2.67(1.04)

Turning our attention to scale version, I have examined if there are differences between gender order versions ("girls and boys" vs "boys and girls") collapsing across single-gender item versions (girls vs boys, from now on referred to as specific gender) and specific gender versions of the scales collapsing across gender order versions. There was a difference in gender order in the full scale, with participants hearing "girls and boys" scoring lower than participants hearing "boys and girls" (16-item scale $t(313) = -2.27$, $p = .03$). However, there was no difference in scores whether participants heard questions about girls or boys on the single-gender items ($p = .20$). There was a difference by gender order ($t(313) = -2.09$, $p = .04$) but not by specific gender for the biological basis component ($p = .11$). There was a difference by gender order ($t(311) = -2.52$, $p = .01$) but not by specific gender for the discreteness component as well ($p = .77$), but there was no difference by gender order or specific gender for the informativeness component ($ps \geq .11$) or the immutability component ($ps \geq .16$; see Table 2-2 for means and standard deviations).

Table 2.2 Means and Standard Deviations for the GES-C and Subscales by Version

	Girl Boy Order	Boy Girl Order	Girl Questions	Boy Questions
16- item Full Scale	2.26 (.57)	2.41(.56)	2.30(.55)	2.37(.58)
Biological Basis	2.39(.77)	2.57(.72)	2.42(.74)	2.55(.76)
Discreteness	2.26(.63)	2.44(.61)	2.34(.59)	2.36(.59)
4-item Informativeness	1.80(.61)	1.85(.60)	1.78(.57)	1.88(.64)
Immutability	2.60(1.04)	2.76(1.05)	2.68(1.08)	2.69(1.02)

Note. Means and Standard Deviations (in parentheses) for each version of the scale.

Comparisons of the GES-C Scores by Participant Gender and Age

Boys scored higher than girls on the full GES-C (16-item $t(313) = -3.09$, $p = .002$), the biological basis component ($t(313) = -2.56$, $p = .01$), the discreteness component ($t(313) = -3.13$, $p = .002$), and the immutability component ($t(313) = -2.38$, $p = .02$). There were no differences by gender for the informativeness components ($p = .34$) or on the Island Task ($p = .19$, see Table 2-3 for means and standard deviations by gender).

Table 2.3 Means and Standard Deviations for the GES-C, Subscales, and Island Task by Gender

	Girls	Boys
16- item Full Scale	2.25(.58)	2.44(.54)
Biological Basis	2.38(.78)	2.60(.70)
Discreteness	2.25(.59)	2.47(.65)
4-item Informativeness	1.80(.56)	1.86(.66)
Immutability	2.55(1.06)	2.83(1.02)
Island Task	10.41(4.17)	11.62(3.06)

Note. Means and standard deviations (in parenthesis) for the GES-C, the components, and the Island Task by participant gender.

Participants' gender essentialism also decreased with age. This was true for the full scale (16-item r = -.41, p < .001), biological basis (r = -.30, p < .001), discreteness (4-item r = -.27, p < .001; 3-item r = -.32, p < .001), informativeness (r = -.26, p < .001), and immutability (r = -.37, p < .001). Participants' scores on the Island Task also decreased with age (r = -.55, p < .001).

Confirmatory Factor Analyses

I first conducted a confirmatory factor analysis (CFA) with no modifications to the 4-factor structure using maximum likelihood estimator in MPlus (Muthén & Muthén, n.d.). The model fit did not meet generally accepted cutoffs of RMSEA ≤ .05, and CFI and TLI ≥ .950 [RMSEA=.095(CI: .085, .105); CFI=.816; TLI=.775; $\chi^2(98) = 375.59$, p < .001] but was a better fit than a 1-factor model [RMSEA=.116(CI: .107, .126); CFI=.708; TLI=.663; $\chi^2(104) = 545.57$, p < .001; $\Delta\chi^2 = 169.98$, p < .001]. The GES-C contains reverse-coded items, which often introduce a method effect into the scale model that is not conceptually relevant (Brown, 2015) and lowers the reliability of a scale (Sliter & Zickar, 2014; Weems & Onwuegbuzie, 2001). To account for the error introduced into the model due to reverse-coded items, I correlated the error of each reverse-coded item with every other reverse-coded item (Brown, 2015). This improved the model fit of the 4-factor CFA [RMSEA=.071 (CI: .059, .083); CFI=.919; TLI=.874; $\chi^2(77) = 199.12$, p < .001]. However, the informativeness item, "Girls/boys all act very different from one another" loaded poorly on the informativeness factor so I conducted a CFA without this item. Model fit was similar to the previous model, but all model items loaded well onto their

respective factors RMSEA=.072 (CI: .059, .084); CFI=.924; TLI=.886; $\chi^2(77) = 180.52$, $p <$ .001; see Table 2.4 for factor loadings for the 16- and 15-item scales].[3]

[3] For means and standard deviations of the 15-item full scale and 3-item informativeness subscale by study sample, version, and participant gender see Appendix B.

Table 2.4 The GES-C Item Loadings for the 4 Factor Confirmatory Factor Analysis

Items	Loadings	
	16-item	**15-item**
Biological Basis		
1. Girls do girl things because they have girl bodies/Boys do boy things because they have boy bodies.	.69	.69
2. You know a kid is a girl or boy/boy or girl because of what their body look like	.55	.55
3. A girl is a girl, even before she is born/A boy is a boy, even before he is born.	.48	.48
4. A kid can choose to be a boy or a girl/boy or a girl; it doesn't matter what their body looks like.	.37	.36
Discreteness		
1. A kid can feel like a girl sometimes and boy other times/A kid can feel like a boy sometimes and girl other times.	.33	.32
2. A kid can feel like a girl and a boy at the same time/ A kid can feel like a boy and a girl at the same time.	.39	.37
3. Girls and boys/boys and girls are the same.	.31	.31
4. Girls and boys/boys and girls are opposites.	.65	.66
Informativeness		
1. Girls all over the world act the same/Boys all over the world act the same.	.39	.39
2. It's easy to tell if a kid is a girl or a boy/boy or girl by how they act.	.49	.49
3. It's easy to tell if a kid is a girl or a boy/boy or a girl by the toys they like.	.90	.90
4. Girls all act very different from one another/Boys all act very different from one another.	.10	-
Immutability		
1. A kid can change from a girl to a boy/boy to a girl.	.73	.73
2. A girl can become a boy if she wants/A boy can become a girl if he wants.	.74	.72
3. If a kid is a girl when they're born, they will be a girl when they grow up, too/If a kid is a boy when they're born, they will be a boy when they grow up, too.	.80	.80
4. A girl can't change that she is a girl/A boy can't change that he is a boy.	.71	.71

Note. CFA latent variable loadings (factor loadings) for the model without and with modifications. (R) indicates a reverse coded item. Generally accepted cut-offs dictate an item loading of .30 or more (Brown, 2015; Worthington & Whittaker, 2006).

Internal Reliability

The reliability statistics (α) for the full scale and components are as follows: .83 (16-item full scale), .84 (15-item full scale), .56 (biological component), .54 (discreteness), .52 (4-item informativeness), .58 (3-item informativeness), and .87 (immutability).

Correlations

The GES-C positively correlated with the Island Task for the 63 participants who completed both measures (r = .49, p < .001; see Table 2-5 for correlations with the GES-C, the components, and the Island Task).

Table 2.5 Correlations between the GES-C, Subscales, and Island Task

	1	2	3	4	5	6	7
1. 16-item Scale							
2. 15-item Scale	.995**						
3. Biological Basis	.82**	.82**					
4. Discreteness	.68**	.69**	.42**				
5. 4-item Informativeness	.57**	.52**	.40**	.21**			
6. 3-item Informativeness	.58**	.57**	.44**	.24**	.93**		
7. Immutability	.84**	.85**	.57**	.46**	.23**	.26**	
8. Island Task	.49**	.49**	.41**	.24	.31*	.35**	.41**

Note. ** indicates p < .001. * indicates p < .05.

General Discussion

To my knowledge, this is the first comprehensive gender essentialism scale for young children that was evaluated for reliability and validity. Although my scale did not meet all

generally accepted model fit criteria (Hu & Bentler, 1999; Sivo et al., 2006), the confirmatory factor analysis revealed fit statistics that are acceptable for this type of measure and sample size (Brown, 2015). Furthermore, the full scale, in both its 16-item or 15-item form, was reliable and correlated with the Island Task (Taylor et al., 2009). Evidencing the validity of the GES-C, the full scale and its components positively correlated with one another and negatively correlated with age, consistent with past research (Rhodes & Gelman, 2009a; Taylor, 1996; Taylor et al., 2009). Yet the reverse-coded informativeness item did not load acceptably onto the informativeness component. Additionally, the subscales of the GES-C would benefit from further, individual validation. Therefore, the GES-C can be beneficial in researching gender essentialism in children; however, it may be best to replace the poorly loading informativeness item and the GES-C components could be further validated.

Creating a scale for young children introduces additional obstacles to scale development that do not present themselves with psychological measures for adults. For example, the sample sizes in the pilot studies and the final study were smaller than what some consider to be ideal. There are many opinions as to what constitutes a valid sample size for a model and what cutoff values for fit indices are ideal (Brown, 2015). In the past, researchers relied on a ratio of number of items to participants to determine sample size. Currently, researchers also use techniques such as Monte Carlo simulations to determine sample size—but these are not without their drawbacks, as a researcher must specify the parameters of the model, which in many cases are unknown. Wolf et al. (2013) found that a CFA of two factors with three indicators all loading at .50 would require 460 participants. The GES-C has four factors with four indicators each and I collected 316 participants. Samples of participants in the hundreds is not typical for experimental developmental studies because of the difficulty in recruiting child participants in comparison to

adult studies that have samples recruited from university subject pools and online platforms with relative ease. Sample size requirements alone make it difficult to conduct several rounds of item testing, and sample size affects model fit statistics (Sivo et al., 2006). Although the fit statistics are not ideal for the GES-C, what constitutes acceptable fit statistics is debated (Brown, 2015) and they seem acceptable and in keeping with other work with children that I am aware of.

For example, consider one recently published psychological measure of social cognition for young children, the Children's Ambivalent Sexism Inventory (CASI; Hammond & Cimpian, 2021). The authors evaluated the scale structure using a confirmatory factor analysis with structural equation modeling in MPlus, and reported similar fit statistics for their two-factor scale as with did with the GES-C [CASI RMSEA = .063; CFI=.903; TLI=.872; SRMR = .051; χ^2 = 65.74 compared to the GES-C RMSEA=.071 (CI: .059, .083); CFI=.919; TLI=.874; SRMR = .062; $\chi^2(77) = 199.12$, $p < .001$]. They also reported similar reliabilities for their two factors ($\alpha s \leq .67$) as those for the four factors of the GES-C ($.52 \leq \alpha s \leq .87$). It is important to note that the CASI has 5-item factors with no reverse-coded items whereas the GES-C has 4-item factors with several reverse-coded items.

Additionally, the components of the GES-C were positively correlated with each other . Although the immutability and informativeness components were weakly correlated, this is consistent with past work that found that components of essentialism did not begin to significantly interrelate for children until the age of 9 years (Gelman et al., 2007). Furthermore, all components are correlated as the theory of essentialism suggests ($.21 \leq rs \leq .57$), supporting the idea of essentialism having multiple components that can be parsed and examined individually and/or examined collectively. Additionally, in the CFA, components were not correlated to the extent that it caused the model to be uninterpretable, a common outcome when

factors are too conceptually similar. Instead, the GES-C demonstrated the observation Rhodes and Mandalaywala (2017) noted in their review, that essentialism is composed of distinct components that are interrelated. The results of the correlations between the GES-C subscales, along with the CFA revealing a better fitting 4-factor than 1-factor model, evidence the multifaceted structure of essentialism.

The components of the GES-C also showed promising reliability. Although only the immutability (α= .87) component exceeded the .60 cutoff for acceptable reliability using Cronbach's alpha, the other components had alphas of .52 and higher. This is impressive given that alphas are known to increase with number of items (Green et al., 1977) and given that reverse-coded items, of which each component had at least one, significantly lower reliability (Sliter & Zickar, 2014; Weems & Onwuegbuzie, 2001). The effects of reverse coding and limited numbers of items may be even larger for young children, given that their data tend to be 'noisier' in general, due to their more limited attention.

Although the Cronbach's alpha scores for the GES-C and its subscales were good, the informativeness item "Girls [boys] all act very different from one another" did not load well onto the informativeness factor. This item was the only reverse-coded item of the component, which could account for some of the poor loading. The component without this item has a Cronbach's alpha of .58 compared to .52 with the item included. Although this item theoretically maps onto the concept of informativeness by implying that gender is uninformative for predicting behaviors and preferences (if all girls are different from one another then gender cannot be used to draw inferences about a girl's behaviors and preferences), it may be the case that stating *all* girls act different from one another seems too strong of a statement to children who may believe that many girls do act the same. Potentially changing this item to "Girls can act very different from

one another" may hedge the statement enough for children to respond to a reverse-coded item similarly to the ways they answer the other items of the informativeness component. These findings and open questions suggest it may be beneficial to test a new reverse coded item for the informativeness component.

I also found a difference in the order in which gender identities were presented. The versions of the scale with items that listed gender identities as "girls and boys" as opposed to "boys and girls" were lower in gender essentialism. It is possible that the order effect is related to the linguistic norm to say "boys and girls" instead of "girls and boys." The standard order of mention may have led participants to report their existing gender essentialist beliefs, whereas the reversed order may have disrupted their default assumptions about gender, thus lowering essentialism. Research has found that the linguistic framing of gender differences can influence gender-based interpretations. Specifically, when men are treated as the linguistic norm by serving as the point of comparison in a statement comparing binary genders (e.g., "Do women lead differently than men?" as opposed to "Do men lead differently than women?"), adults are more likely to endorse gender stereotypes and accept power differences between genders as legitimate (Bruckmüller et al., 2012). Similarly, when boys are treated as the linguistic norm in comparative statements (e.g., "Girls are as good as boys at math"), children assume that boys are probably better (Chestnut et al., 2021). As a result, it may be useful to choose one version of the scale to use with all participants in a particular study to avoid any differences in essentialism due to gender order. Future scale testing could confirm this difference due to wording and, if confirmed, identify mechanisms that account for this difference.

Beyond the statistical examination of the coherence of the GES-C, I assured my scale was theoretically relevant to gender essentialism. The components I tested were adapted from

existing adult measures that inspired some of the GES-C items (Bastian & Haslam, 2006; S. R. Levy et al., 1998). Furthermore, the GES-C correlated with another well-known gender essentialism measure, the Island Task (Taylor et al., 2009). The Island Task, at face value, measures the biological and informativeness components of gender essentialism, however it was not created to measure biological basis and informativeness separately and therefore conflates essentialist components. The GES-C full scale, and the biological basis, informativeness, and immutability components positively correlated with the Island Task. Given the GES-C discreteness subscale did not significantly correlate with the Island Task and that a 4-factor CFA model was the best fit for the data, it may be that the GES-C is measuring with enough specificity to be able to analyze the components separately in future studies. To this point, it would be advantageous to further validate the GES-C component subscales independently. Future work could compare the GES-C components to existing measures of the individual measures. For example, the immutability component scores could be compared to a version of the gender constancy scale by Slaby and Frey (1975) or the immutability task in which a participant views a picture of a child and chooses a picture of an adult that best represents what the child looks like in the future (Kinzler & Dautel, 2012; Pauker et al., 2010, 2016; Roberts & Gelman, 2016; Ruble et al., 2007). One could also consider examining outcomes specific to a particular component of gender essentialism. For example, beliefs that gender is informative should relate to endorsement of gender stereotypes. Overall, these correlations promising start to establishing the validity of the GES-C.

Also, the full scale and all components showed the same patterns of relation to participant characteristics reported in extant research, pointing to the validity of the GES-C. Certain demographics have been found to relate to more or less endorsement of gender essentialism. For

example, boys had higher scores on the GES-C than girls, consistent with past work finding that college-aged men endorsed gender essentialism more than college aged women (Smiler & Gelman, 2008). Also, GES-C scores declined with age, consistent with past research on gender essentialism using different measures (Rhodes & Gelman, 2009; Ruble et al., 2007; Taylor, 1996; Taylor et al., 2009). Both of these relations with participant characteristics should be expected in a measure of gender essentialism and point to the validity of the GES-C.

The GES-C, because it was based on theory, may also be able to be adapted to examine essentialism in other social categories. There is a large developmental literature of essentialism related to race, religion, and nationality (Deeb et al., 2011; Diesendruck & Haber, 2009; Diesendruck & Menahem, 2015; Heiphetz et al., 2017; Hussak & Cimpian, 2019; Pauker et al., 2010; Smyth et al., 2017). Items like "It's easy to tell if a kid is a boy or a girl by how they act" could be adapted to look at nationality. For example, the item could read "It's easy to tell if a kid is French or American by how they act." It may be beneficial to adapt this measure to other social categories and evaluate the validity and reliability of the scale given the research that shows interesting differences and similarities in the ways adults and children essentialize different social categories (Diesendruck et al., 2013; Haslam et al., 2000; Rhodes et al., 2012; Rhodes & Gelman, 2009; Roberts & Gelman, 2016).

One limitation of the GES-C is that it uses specific gender identities. The items in the GES-C require comparing one gender identity to another instead of asking about all gender identities. I used the terms boy and girl. Existing gender essentialism measures focus on the binary genders of girl/women and boy/man as well. However, in the future there is the opportunity to use nonbinary and gender nonconforming identities in the GES-C items. It may be

informative to see if nonbinary gender identities elicit different levels of gender essentialist beliefs.

The GES-C offers developmental researchers a more comprehensive and efficient way to measure gender essentialism. Researchers can now measure a broader range of gender essentialism beliefs within a single measure. Given the broader range of beliefs, it would be beneficial to validate the components of the GES-C. It may also be advantageous to replace the reverse-coded informativeness item of the GES-C as it did not load well onto the informativeness factor. Although one item of informativeness may need to be replaced, the full GES-C is a promising new measurement tool to assess gender essentialism in children.

Chapter 3 Indirect Contact with Transgender Identities through Stories and its Influences on Children's Gender Essentialism

Recently, people with diverse gender identities have garnered more attention in the media, politics, and society in general (David, 2017; Horak, 2014; Steinmetz, 2014). Given this relatively large increase in the visibility of trans people, the moment is ripe for better understanding how gender diversity may influence children's conceptualizations of gender. Gender identities outside of the binary may challenge adults to rethink their notions about gender, including their essentialist beliefs (Fine et al., n.d.). For example, binary transgender people may challenge biological basis beliefs because they have a gender that is different from the gender their body suggests. They may also challenge informativeness beliefs because people may not know what gendered behavior to expect if they know a person is trans. As a result, this may blur the boundaries between girl and boy for some. And finally, a transgender person's transition may be evidence to some of the malleability of gender.

The present study examines whether introducing young children to gender diversity would likewise influence endorsement of gender essentialist beliefs, at a time in development when these beliefs are already in flux (Rhodes & Gelman, 2009a; Taylor, 1996; Taylor et al., 2009). However, explaining transgender identities (identities of people whose gender is a different binary gender than the gender they were assigned at birth) to young children may be difficult, given children's more strict understanding of gender at young ages (Trautner et al.,

2005). Even adults struggle with how to understand and accept transgender people (Elischberger et al., 2016; Payne & Smith, 2014).

One potential mode of explaining diverse gender identities to children is stories. Recently, children's books have been written that aim to help young children understand what it means to be transgender. This literature may be instrumental in exposing children to transgender identities and influencing children's gender essentialist beliefs. Yet, because of the discrimination and prejudice transgender people face (Kosciw et al., 2018; Miller & Grollman, 2015), it is important that books convey their messages in a way that is sensitive and informative. And, although there is evidence that children learn from stories, whether stories are a good means of delivering messages about the social structure of the world has been less studied (Strouse et al., 2018). Furthermore, how closely stories reflect real-life may greatly impact how well knowledge is transferred to children's everyday lives (Richert & Smith, 2011).

With this second study, I examine how a realistic story with a transgender girl and a fantastical story with anthropomorphized marker with an outside that does not match his inside explain trans identities to young children, and, importantly, how these stories influence children's endorsement of gender essentialist beliefs. I look specifically at children in two age groups, five- to six-year-olds and nine- to ten-year-olds, because prior research shows developmental changes at these ages in endorsement of gender essentialism (Rhodes & Gelman, 2009a; Taylor, 1996; Taylor et al., 2009). The creation of my scale in Study 1 affords me the ability to analyze the influence of these stories at a more macro level with a comprehensive measure of gender essentialism, and the more granular level of the four components of gender essentialism: biological-basis, discreteness, informativeness, and immutability.

Impact of Intergroup Contact on Essentialist Beliefs

One way for children to learn about trans identities is to have contact with trans people. Intergroup contact has been studied in adults and children, and found to reduce biases and discrimination against outgroup members such as people of other races, ethnicities, sexual orientations, and abilities (Aboud & Brown, 2012; Guerra et al., 2010; Pettigrew & Tropp, 2006). Allport (1954) proposed that contact with stigmatized outgroup members can lead to reduced biases, especially under optimal conditions. Importantly, some recent research also indicates that intergroup contact relates to changing essentialist beliefs (Deeb et al., 2011; Fine et al., n.d.; Lytle et al., 2017b; Pauker et al., 2017).

Although there is research to suggest that contact can influence essentialist beliefs, the work is nascent, especially in the developmental literature. I know of two studies that have looked at the relation between essentialism and contact for children. However, both studies look at race and ethnicity and are correlational. Essentialist beliefs about ethnicity were found to be lower in Israeli and Arab children who attended a multi-ethnic school in Israel compared to mono-ethnic schools (Deeb et al., 2011). Pauker et al. (2016) also found that older children of their four- to eleven-year-old participants differed in race essentialism as a function of the state in which they resided. Older children from Massachusetts, a less racially diverse state, expressed more race essentialism and race stereotyping than older children from Hawaii, a more racially diverse state. So, there is some evidence to believe that cisgender children's gender essentialist beliefs may be affected by contact with transgender children, but the field is in need of more research around the relation between essentialism and contact generally, as well as specifically related to gender.

However, researchers estimate that the population of trans people is very small, making it less likely that children will meet many trans people in their day-to-day life. A recent meta-

analysis led researchers to estimate the trans population to be 0.39% (Meerwijk & Sevelius, 2017). This presents a problem in encouraging the type of direct, in-person contact Allport (1954) originally proposed. Importantly, however, the intergroup contact literature has greatly expanded since Allport first proposed his theory, to encompass indirect contact including contact with media that features outgroup members (Hoffarth & Hodson, 2018; Tropp et al., 2016). Indirect contact has also been shown to reduce biases and anxiety about outgroup members in adults and children (Cao & Meng, 2020; Husnu et al., 2018; Pettigrew et al., 2007; Vezzali et al., 2017; Wölfer et al., 2019), raising the question of whether indirect contact might also change essentialist beliefs. Indirect contact includes media, such as news and social media posts. And, these types of media contact involving LGBTQ people have been found to relate to positive attitudes towards the LGBTQ community in adults (Lissitsa & Kushnirovich, 2020). Given that many children may not have the opportunity to have in-person interactions with trans people, it is important to find indirect means of contact with trans people. Children's stories, a form of media, offer a way to provide children with indirect contact.

Children's Ability to Learn from Media

Children's literature and TV programming have long been used to educate children on social issues, academic subjects, and morality. And research supports the idea that children can learn from stories under certain conditions (for a review see Strouse et al., 2018). Much research on children's media, both books and TV shows, focuses on academic themes such as science, analogical problem solving (i.e., transferring a solution from one context to another similar context), and vocabulary (e.g., Ganea et al., 2008; Kelemen et al., 2014; Richert et al., 2009). Although less is known about how children learn from stories that teach social lessons about culture, diversity, and morality, children can recognize the morals conveyed through stories

(Larsen et al., 2018; Walker & Lombrozo, 2017). Children as young as four years old who heard a story about a child sharing were more likely afterwards to share stickers compared to a control group (Larsen et al., 2018). Walker and Lombrozo (2017) found when children as young as five years old were read a story and asked to explain the feelings of the main characters, they were able to glean the moral. Children's TV programming also aims to teach children about cultural, racial, and ethnic diversity. A meta-analysis of the impact of Sesame Street, a show known for highlighting and promoting interracial and interethnic relationships, on young children's educational outcomes finds the show does promote academic and social learning (Mares et al., 2015). All this work suggests children's media may be an impactful way to educate children on important real-world lessons.

Storytelling via books and shows can be a powerful way to teach young children; however, these media often feature fantastical characters and situations. Fantasy, depicting unrealistic protagonists and scenarios, plays a big role in children's learning and transferring information from the story to the real world. There is evidence that even very young children can separate the real world and fantasy (Woolley, 1997), and these boundaries between real life and fantasy may make fantastical stories less effective for teaching children about real life concepts (Richert & Smith, 2011; Sobel et al., 2013; Walker et al., 2015; Woolley, 1997; Woolley & Ghossainy, 2013). On the one hand, preschool children can learn biology lessons from anthropomorphized animals in storybooks while still realizing that animals do not possess the same psychological properties as humans (Geerdts et al., 2016). On the other hand, three and half to six year old children were better at solving analogical problem solving tasks after hearing a story with real people than after hearing a story with fantastical characters (Richert et al., 2009). Furthermore, Larsen and colleagues (2018) found that when four- to six-year-old children were

told a story about sharing, with a raccoon as the protagonist, they were more selfish immediately following the story. When children in another condition of the same study were told a story about sharing, with a human protagonist, they shared more afterwards. And, in contrast to the meta-analysis of Sesame Street noted above, animated shows not including Sesame Street that feature protagonists of color do little if anything to teach children about other cultures or encourage positive feelings towards children of color (Mares et al., 2015).

Richert and Smith (2011) offer a theory of quarantining to explain why children often do not apply messages from fantastical stories to real life. They maintain that children quarantine the relevant information that they glean from fantastical stories and do not transfer that information to real life as easily as they do with realistic stories, because they know parts of fantastical stories are fantasy but are still unclear as to which parts of the fantastical stories could be real. The researchers suggest children assert strong boundaries between fantasy and real life as they continue to learn what is possible in the real world. Walker and colleagues' (2015) work supports this theory, showing that three- to six-year-old children make causal inferences from realistic stories more often than from fantastical stories. So, while children do learn from stories, both fantastical and realistic, the features of the story and the type of information the story conveys are important in determining how successfully children transfer the knowledge presented to them.

The Present Study

In this study, I explored how contact with trans identities through stories impacted children's gender essentialist beliefs. The stories I developed were based on two books that address trans identities, *I am Jazz* (Herthel & Jennings, 2014) and *Red: A crayon's story* (Hall, 2015). *I am Jazz* tells Jazz Jennings's autobiographical story of transitioning from a boy to a girl

in early childhood. In *Red: A crayon's story*, Red is an anthropomorphized blue crayon with a red label. He is also treated as a red marker by those around him. By the end of the story, Red transitions to Blue. This story is not about gender, but may be considered analogous to trans identities because inside the protagonist is blue, but his outside is red. Both stories emphasize how their protagonists' outsides do not match how they feel on the inside. The stories emphasize how the physical self is separate from one's felt identity, and show how the characters around the protagonists (e.g., friends, family) grow to understand and support the protagonists' trans identities.

I examined if children learned about transgender identities from these two stories (one direct, one metaphorical), and whether these stories affected children's gender essentialism, as compared to children in a control group who did not hear a story. I predicted that children who heard these stories would be more likely to say a child can be transgender and that this would affect their gender essentialist beliefs when compared to children in the control condition.

Additionally, I explored if understanding transgender identities and endorsement of gender essentialism were influenced by whether children heard the direct (Jazz) or metaphorical (Blue) story. Young children especially may find the metaphorical story difficult to translate into a lesson about gender identity and therefore not benefit as much as they do from the direct story in terms of understanding trans identities (Richert & Smith, 2011). This may also result in children in the metaphorical Blue condition not differing from children in the control condition on gender essentialism. I also examined if there were differences in the relation between watching a story and gender essentialism by age, as past work found that children's gender essentialism can lessen with age, with a noticeable difference between five- and six-year-old children and nine- to ten-year-old children, with older children attributing gender-related

behavior more to environmental reasons and mechanisms (Rhodes & Gelman, 2009a; Taylor, 1996; Taylor et al., 2009). Because children struggle to translate lessons from fantastical stories at five to six years of age (Larsen et al., 2018; Walker et al., 2015), and children of this age exhibit more gender essentialism than nine- to ten-year old-children, the younger children may show a greater difference between the two stories than older children.

Finally, I wanted to investigate how my two stories influenced the biological basis, informativeness, discreteness, and immutability components of gender essentialism individually. As I mentioned earlier, it is still an open question how contact with a transgender character may influence the components of gender essentialism. As a result, I am interested in how participants' scores for each component are influenced by each story compared to the control condition. I examined age differences in component scores by condition. Very little work has addressed the impact of media, both fantastical and direct, on children's learning about the social learning, so it is important not only to evaluate gender essentialism in general but also the components separately, as these components are not always equally endorsed within a category (Gelman, 2004; Rhodes & Mandalaywala, 2017).

Method

Participants

I recruited 173 participants from my desired age groups: 90 5–6-year-old participants (M = 5.44 years, SD = .50) and 83 9-10-year-old participant (M = 9.31 years, SD = .47). All participants were cisgender as reported by their parents. Ninety-four (54.3%) participants were girls and 79 (45.7%) were boys. Eighty-four (48.6%) children were White, 21 (12.1%) were multiracial, 8 (4.6%) were Asian, 4 (2.3%) were Latinx, 4 (2.3%) did not report race, and 52

(30%) were not asked due to an error when transitioning the consent form to an online format. Seventy-eight (45.1%) participants completed our study in-person in our lab. Due to the COVID-19 pandemic, 95 (54.9%) of our participants completed our study online via secure Zoom meetings. Participants from this study were also used in Study 1. No participants were excluded.

Materials

With the help of two research assistants, I created a video version of each story. Each video was voiced by the same person, and the same artist drew images during the video to accompany the text to keep participants engaged. The artist's hand was visible while she was drawing in both videos. Both videos were 3:33 minutes long. The Jazz story video used a subset of the text from *I am Jazz* and featured drawings based on the illustrations from the book. The Marker story video adapted the text from *I am Jazz* to be directly parallel, and relevant to activities a marker would do. We used a marker so that we could put blue ink into the plastic shell of a red marker that the artist drew with, in the video, and the inside ink would be invisible, just like gender identity is for people.

Below is the script for both clips.

Table 3.1 Story Condition Scripts

Jazz Story Script	Marker Story Script
I am Jazz! For as long as I can remember, my favorite color has been pink. My second-favorite color is silver and my third-favorite color is green.	I am Blue! For as long as I can remember, my thing to draw is a blue whale. My second-favorite thing to draw is blue jeans and my third-favorite thing to draw is blue suede shoes.

Here are some of my other favorite things: dancing, singing, back flips, drawing, soccer, swimming, makeup, and pretending I'm a pop star. Most of all, I love mermaids. Sometimes I even wear a mermaid tail in the pool!

Here are some of my favorite things to draw: blueberries, oceans, skies, bluebirds, and bluebells. Most of all, I love drawing blue whales! Sometimes I even draw blue whales in a blue ocean!

My best friends are Samantha and Casey. We always have fun together. We like high heels and princess gowns, or cartwheels and trampolines. But I'm not exactly like Samantha and Casey.

My best friends are Royal and Sky. We always have fun together. We like drawing blue things together. But I'm not exactly like Royal and Sky.

I have a girl brain but a boy body. I was born this way!

I have a blue inside but a red outside. I was made this way!

Clip pauses for experimenter to ask a question to scaffold understanding and give answer.

Experimenter: "What is different about Jazz?"

Experimenter: "What is different about Blue?"

Participant's answer is recorded.

Experimenter: "Jazz has the body of a boy but is a girl on the inside."

Experimenter: "Blue has the body of a red marker but is a blue marker on the inside."

When I was very little, and my mom would say, "You're such a good boy," I would say, "No, Mama. Good GIRL!" At first my family was confused. They'd always thought of me as a boy.

When I was very little, and my mom would say, "You're such a good red marker," I would say, "No, Mama. Good BLUE MARKER!" At first my family was confused. They'd always thought of me as red.

As I got a little older, I hardly ever played with trucks or tools or superheroes. Only princesses and mermaid costumes. My brothers told me this was girl stuff. I kept right on playing. My sister says I was always talking to her about my girl thoughts, and my girl dreams, and how one day I would be a beauuuutiful lady. She would giggle and say, “You’re a funny kid.”

Then one amazing day, everything changed. Mom and Dad took me to meet a new doctor who asked me lots and lots of questions. That night at bedtime, my parents both hugged me and said, “We understand now. Be who you are. We love you no matter what.” This made me smile and smile and smile.

Mom and Dad told me I could start wearing girl clothes to school and growing my hair long. They even let me change my name to Jazz. Being JAZZ felt much more like being ME! Mom said that being Jazz would make me different from the other kids at school, but that being different is okay. What’s important, she said, is that I’m happy with who I am.

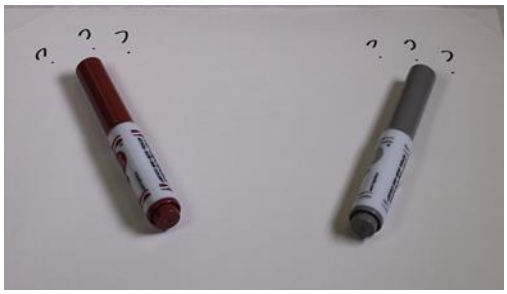
Being Jazz caused some other people to be confused too, like the teachers at school. At the beginning of the year they wanted me to use the boys’ bathroom, and play on the boys’ team in gym class, but that didn’t feel normal to me at ALL.

As I got a little older, I hardly ever drew stop signs or fire trucks. Only bluebirds and blueberries. My brother told me this was blue stuff. I kept right on drawing. My sister says I was always talking to her about my blue thoughts, and my blue dreams, and how one day I would be a beauuuutiful blue marker. She would giggle and say, “You’re a funny kid.”

Then one amazing day, everything changed. Mom and Dad took me to meet a new doctor who asked me lots and lots of questions. That night at bedtime, my parents both hugged me and said, “We understand now. Be who you are. We love you no matter what.” This made me smile and smile and smile.

Mom and Dad told me I could start drawing blue things at school. They even let me change my name to Blue. Being BLUE felt much more like being ME! Mom said that being Blue would make me different from the other kids at school, but that being different is okay. What’s important, she said, is that I’m happy with who I am.

Being Blue caused some other people to be confused too, like the teachers at school. At the beginning of the year they wanted me to draw red things, and play on the red team in gym class, but that didn’t feel normal to me at ALL.



Clip pauses for experimenter to ask a question to scaffold understanding and give answer.

Experimenter: "Why is Jazz sad?"

Experimenter: "Why is Blue sad?"

Participant's answer is recorded.

Experimenter: "Jazz is sad because people treat her like a boy."

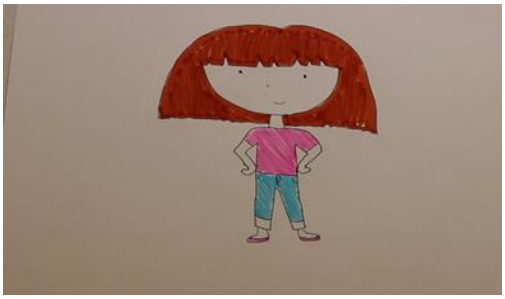
Experimenter: "Blue is sad because people treat him like a red marker."

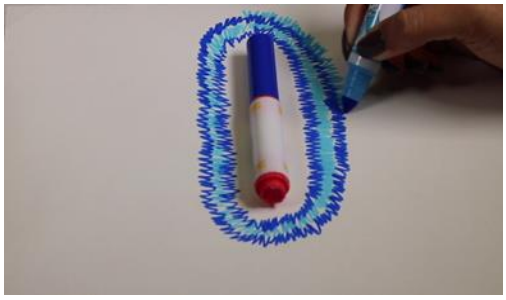
I was so happy when the teachers changed their minds. I can't imagine not playing on the same team as Casey and Samantha. Even today, there are kids who tease me, or call me by a boy name, or ignore me altogether. This makes me feel crummy. Then I remember that the kids who get to know me usually want to be my friend. The say I'm one of the nicest girls at school.

I don't mind being different. Different is special! I think what matters most is what a person is like inside. And inside, I am happy. I am having fun. I am proud! I am Jazz!

I was so happy when the teachers changed their minds. I can't imagine not playing on the same team as Royal and Sky. Even today, there are kids who tease me, or call me by a red name, or ignore me altogether. This makes me feel crummy. Then I remember that the kids who get to know me usually want to be my friend. The say I'm one of the nicest blue markers at school.

I don't mind being different. Different is special! I think what matters most is what a person is like inside. And inside, I am happy. I am having fun. I am proud! I am Blue!





Clip ends and experimenter asks a question to scaffold understanding and give answer.

Experimenter: "Why is Jazz happy?"

Experimenter: "Why is Blue happy?"

Participant's answer is recorded.

Experimenter: "Jazz is happy because people treat her like a girl."

Experimenter: "Blue is happy because people treat him like a blue marker."

After the video, participants are asked, "What did you learn from the story?" Their responses are recorded but no answer is given.

Measures

We used the same measures from Study 1, the Gender Essentialism Scale for Children (GES-C Full Scale α = .82, Biological Basis α = .50, Discreteness α = .44, Informativeness α = .58, Immutability α = .82) and the Island Task (Taylor et al., 2009).

Transgender Identity Understanding. All participants, regardless of condition, were asked "Can a girl be a boy on the inside? Yes, maybe, or no." and "Can a boy be a girl on the inside? Yes, maybe, or no." Participants who heard Blue's story were additionally asked, "Can an orange marker be green on the inside? Yes, maybe, or no." and "Can a green marker be an orange marker on the inside? Yes, maybe, or no." To analyze responses, yes was coded as "3," maybe as "2," and no as "1." After coding the responses, the two gender questions were averaged together. The same process was followed for the marker questions. Final scores ranged from 1-3, with 3 indicating the most understanding of transgender identity.

Island task (Taylor et al., 2009). This task is the same task used in Studies 1 and 2. Participants were told a story about a child who lives on an island with people of the binary different gender from the child's gender. Then they answered questions about the preferences, activities, gender-based roles, and biology of the child. Scores ranged from 1 to 15, with higher scores indicating more essentialism.

Procedure

Participants were randomly assigned to one of three conditions: watching the story video of Jazz, watching the story video of Blue, and completing the measures without watching a story video (the control condition). The videos included three comprehension questions at different

points through the video (see Table 3.1), including: “What is different about Jazz/Blue?”, “Why is Jazz/Blue sad?”, and “Why is Jazz/Blue happy?” Regardless of the answer the child gave to these questions they were told, “Jazz/Blue has the body of a boy/red marker but is a girl/blue marker on the inside,” “Jazz/Blue is sad because people treat her/him like a boy/red marker,” and “Jazz/Blue is happy because people treat her/him like a girl/blue marker.” Afterward, all children in the video conditions were asked, “What did you learn from the story?”

If a participant watched the Blue clip, they answered the transgender identity understanding questions about marker colors immediately after being asked about the meaning of the story. All participants, including those in the control condition, answered the questions about transgender identity understanding about girls and boys. Next, participants completed the GES-C and then completed the Island Task.

Results

I have organized the results by preliminary and main analyses addressing my hypotheses and exploratory questions described in the Present Study section. These hypotheses were that participants who heard a story would score higher on transgender understanding questions than those in the control condition, and that participants who heard a story would score lower on gender essentialism measures than those in the control condition. Additionally, I examined differences in transgender understanding and essentialism measures between condition by age, because of past work finding that children’s gender essentialism lessens with age (Rhodes & Gelman, 2009a; Taylor, 1996; Taylor et al., 2009) and that children’s ability to glean messages from fantastical stories increases with age (Larsen et al., 2018; Walker et al., 2015). I also examined the effect of condition on each component of the GES-C overall, and by age group.

Preliminary Analyses

The preliminary analyses examined whether there were any differences in our sample by the manner in which participants completed the study (i.e., in-lab vs. online) and questions we asked of parents. These questions asked parents if they had heard of the books *I am Jazz* or *Red: A crayon's story* and if their child knew a person who is gender nonconforming. Finally, I describe any age differences regardless of condition, as past research has found that essentialism is lower in older children.

Prior to conducting my main analyses, I looked to see if there were differences on the gender essentialism scale, the Island Task, or the transgender understanding question between the in-lab and online participants and found none ($p \geq .18$). I also found no difference between in-lab and online participants' responses on the individual components of essentialism ($p \geq .12$).

I also asked parents of participants if they had heard of either of the books that inspired this study; 57 (32.9%) had heard of at least one book, 80 (46.2%) had not heard of the books, and 36 (20.8%) were not asked the question as it was added after we began data collection. Of the 137 parents who were asked the question, 41.6% had heard of at least one book and 58.4% had not heard of either book. Parents' knowledge of the books did not correspond to participants' responses on any measure, including the components (ps $\geq .14$; see Table 3.2 for distribution of answers to this question by age and condition).

Table 3.2 Number of participants by age whose parents had heard of *I am Jazz* and *Red: A crayon's story*

	Condition	5-6 yr olds	9-10 yr olds
Yes	Jazz	13	10
	Marker	9	7
	Control	6	12
	Total	28	29
No	Jazz	10	13
	Marker	12	18
	Control	16	11
	Total	38	42
Not Asked/No Response	Jazz	6	4
	Marker	8	4
	Control	10	4
	Total	24	12

Twenty-nine parents (16.8%) reported their child knowing a gender-nonconforming person, 107 (68.8%) reported their child not knowing anyone who is gender non-conforming, and 37 (21.4%) did not respond. Of those 136 who answered the question, 21.3% answered their child knew someone, and 78.7% answered their child did not know someone. Participants who knew someone who is transgender or gender nonconforming (i.e., a person whose gender does not coincide in part or in whole, with the norms of the gender/sex assigned to them at birth; GNC) were significantly less gender essentialist in general (GES-C Yes M = 1.77, SD = .48, No M = 2.35, SD = .54, $t(132) = -5.12$, $p < .001$; Island Task Yes M = 8.89, SD = 3.98, No M = 11.90, SD = 3.29, $t(1320 = -4.09$, $p < .001$), reported significantly less essentialism within each component except discreteness (Biological basis Yes M = 1.89, SD = .74, No M = 2.46, SD = .66, $t(132) = -3.93$, $p < .001$; Discreteness Yes M = 2.06, SD = .63, No M = 2.30, SD = .65, $t(132) = -1.68$, $p = .10$; Informativeness Yes M = 1.44, SD = .37, No M = 2.03, SD = .71, $t(89.59) = -6.28$, $p < .001$; Immutability Yes M = 1.69, SD = .85, No M = 2.61, SD = .65, $t(132)$

= -4.82, $p < .001$), and demonstrated more understanding of transgender identities (Yes M = 2.86, SD = .52, No M = 2.49, SD = .68, $t(53.05) = 3.09$, $p = .003$; see Table 3.3 for distribution of answers to this question by age and condition).[4],[5]

Table 3.3 Number of participants by age whose parent reported participant knew a gender-nonconforming person

	Condition	5-6 yr olds	9-10 yr olds
Yes	Jazz	0	10
	Marker	3	9
	Control	3	4
	Total	6	23
No	Jazz	23	13
	Marker	18	16
	Control	19	18
	Total	60	47
No Response	Jazz	6	4
	Marker	8	4
	Control	10	5
	Total	24	13

Age Differences in Essentialism

Consistent with past literature (Taylor et al., 2009), older children were less gender essentialist than younger children, on both the 16-item scale (5-6-year-olds M = 2.58, SD = .49; 9-10-year-olds M = 1.90, SD = .47; $t(169) = 9.25$, $p < .001$) and the Island task (5-6-year-olds M = 12.71, SD = 2.95; 9-10-year-olds M = 9.54, SD = 3.82; $t(152.35) = 5.97$, $p < .001$). That pattern held for all components of the GES-C (ps < .001; see Table 3.4 for means and standard

[4] Any t test in this results section with degrees of freedom that are not whole numbers indicates I could not assume equal variances.

[5] Twenty-three participants who knew a GNC person were older children, and 6 participants were younger children. However, even excluding these 29 participants, older children were significantly less gender essentialist on all measures and were more understanding of transgender identities than younger children (ps < .001).

deviations of measures by age). Furthermore, older children demonstrated a better understanding of transgender identities than younger children (5-6-year-olds $M = 2.25$, $SD = .79$; 9-10-year-olds $M = 2.85$, $SD = .37$; $t(129.87) = -6.45$, $p < .001$).

Table 3.4 Means and standard deviations for each measure by age group and the p-value for the difference between groups

Measures	*Age Group*		*p*
	5-6 yr olds	9-10 yr olds	
GES-C	2.58(.49)	1.90(.47)	<.001
Biological basis	2.70(.69)	2.02(.66)	<.001
Discreteness	2.43(.74)	1.98(.50)	<.001
Informativeness	2.21(.75)	1.61(.51)	<.001
Immutability	2.96(.89)	1.99(.87)	<.001
Island Task	12.71(2.95)	9.54(3.82)	<.001
Transgender understanding	2.25(.79)	2.85(.37)	<.001

Main Analyses

The main analyses are divided into sections. In the first section, I report an analysis of variance that tested if there was an interaction between age and condition. I examined differences in understanding trans identities by condition and age group. The next section reports differences on the GES-C full scale and component scores by condition and age group. Finally, I examined differences on the Island Task by condition and age group.

Age, Condition, and Age by Condition Interaction by Measure

I conducted analyses of variance (ANOVAs) to examine if there were significant main effects of condition and age, and if there were interactions of age by condition for all measures. The main effect of age was significant for all measures. Condition was significant for the immutability component and the transgender understanding component (see Table 3.5 for *F* statistics). Although none of the interactions between age and condition reached significance, I will analyze differences between conditions for each age group because they were planned and omnibus *F* tests can obscure significant differences between independent variables (Rosenthal & Rosnow, 1985).

Table 3.5 Results from ANOVAs of Condition, Age, and Age by Condition for all measures

Measures	Age				Condition				Age x Condition			
	F	*df*	*p*	η^2	F	*df*	*p*	η^2	F	*df*	*p*	η^2
GEM-C	**83.99**	**1, 165**	**< .001**	**.34**	1.55	2, 165	.22	.02	.01	2, 165	.99	.00
Biological basis	**43.28**	**1, 165**	**< .001**	**.21**	1.84	2, 165	.16	.02	.18	2, 165	.83	.00
Discreteness	**21.86**	**1, 165**	**< .001**	**.12**	1.65	2, 165	.19	.02	2.03	2, 165	.13	.02
Informativeness	**37.52**	**1, 165**	**< .001**	**.19**	1.67	2, 165	.19	.02	.35	2, 165	.73	.00
Immutability	**51.77**	**1, 165**	**< .001**	**.24**	**4.43**	**2, 165**	**.01**	**.05**	.24	2, 165	.79	.00
Island Task	**34.23**	**1, 165**	**< .001**	**.18**	.64	2, 165	.53	.01	.53	2, 165	.59	.01
Transgender understanding	**41.24**	**1, 165**	**< .001**	**.20**	**5.95**	**2, 165**	**.003**	**.07**	2.41	2, 165	.09	.03

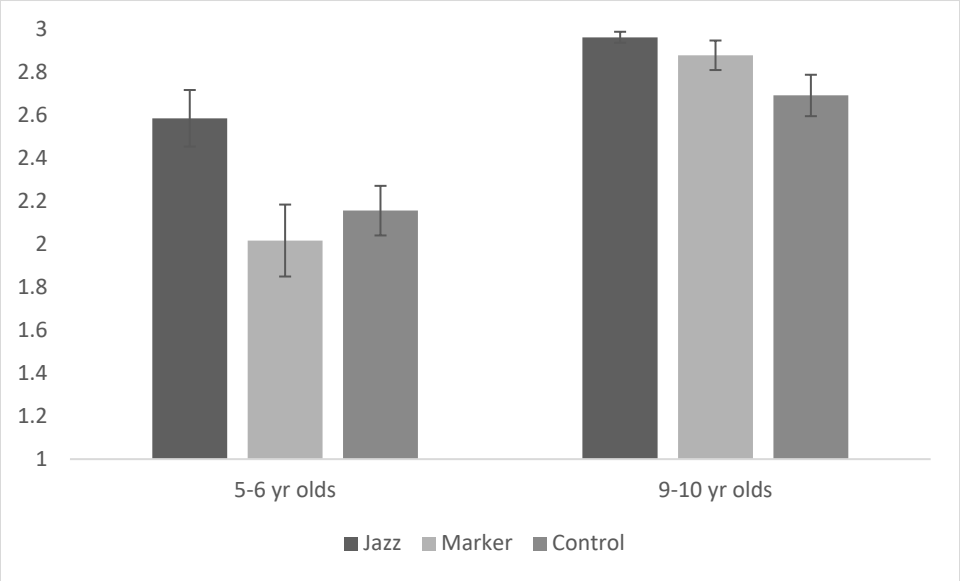
Understanding of Trans Identities by Condition and Age

I examined if children who heard a story displayed a better understanding that a transgender person feels like a gender different from the sex assigned to them at birth. On average, regardless of condition, participants seemed to acknowledge to some degree that a person could feel like a gender different from the gender assigned to them based on their body as a one sample t-test revealed all averages of understanding of transgender identities were significantly higher than the midpoint (*ps* < .001). Looking at all participants, those in the combined story conditions (*M* = 2.61, *SD* = .71) were not significantly different from those in the control condition (*M* = 2.40, *SD* = .64; Stories vs. Control *t*(170) = 1.89, *p* = .06). However, participants in the Jazz condition (*M* = 2.77, *SD* = .55) better understood what it means to be transgender than those in the Marker condition (*M* = 2.45, *SD* = .81; Jazz vs. Marker *t*(100.44) = -2.48, *p* = .02) and the Control condition (*M* = 2.40, *SD* = .64; Jazz vs. Control *t*(110.39) = -3.33, *p* = .001). Those in the Marker condition did not significantly differ from those in the control condition (Marker vs. Control *t*(108.27) = .38, *p* = .70).

Looking at younger children, I found the same pattern of results. The combined story conditions (*M* = 2.30, *SD* = .85) did not differ from the control condition (*M* = 2.16, *SD* = .65; Stories vs. Control *t*(78.84) = -.90, *p* = .37). The Jazz condition (*M* = 2.59, *SD* = .71) improved understanding of transgender identities compared to the Marker condition (*M* = 2.02, *SD* = .90; Jazz vs. Marker *t*(53.03) = -2.67, *p* = .01) and the Control condition (Jazz vs. Control *t*(59) = 2.46, *p* = .02). The Marker condition did not significantly differ from the Control condition (Marker vs. Control *t*(50.63) = -.68, *p* = .50).

In contrast, for older children, the composite score including both stories (*M* = 2.92, *SD* = .26) boosted understanding compared to the control condition (*M* = 2.69, *SD* = .49; Stories *vs.* Control *t*(32.28) = 2.22, *p* = .03). The Jazz story (*M* = 2.96, *SD* = .13) promoted more understanding compared to the control condition (Jazz vs. Control *t*(28.54) = 2.71, *p* = .01). However, the Marker story (*M* = 2.88, *SD* = .37) did not promote more significantly understanding compared to the control condition (Marker vs. Control *t*(108.27) = 1.58, *p* = .12). But the Jazz story and the Marker story did not significantly differ in the amount of understanding they promoted (Jazz vs. Marker *t*(35.63) = 1.14, *p* = .26; see Figure 3.1).

Figure 3.1 Understanding of Transgender Identities by Condition and Age Group



Note. Higher numbers signify more understanding, on a scale of 1-3. Error bars represent standard errors for conditions and age category.

Further examining how the Marker story affected understanding of trans identity, I looked at the questions of trans understanding of marker color and tested whether those answers differed from the transgender understanding answers participants gave. The intention of the

marker trans understanding of gender questions was to help participants make the connection between the story's metaphor of marker color to gender identity by presenting questions about marker color in the same way in which later gender questions were asked. However, these questions could also serve as a measurement of if children inferred the marker story was a metaphor for gender by analyzing how similar answers were on the color and gender questions. Overall, participants in the Marker condition did not differ in their response regarding marker color versus transgender understanding (marker color $M = 2.50$, $SD = .69$; transgender $M = 2.45$, $SD = .81$; $t(57) = .62$, $p = .54$). Younger children did not give significantly different answers (marker color $M = 2.10$, $SD = .75$; transgender $M = 2.02$, $SD = .90$; $t(28) = .61$, $p = .55$). Older children's answers did not significantly differ from one another (color $M = 2.90$, $SD = .31$; transgender $M = 2.88$, $SD = .37$; $t(28) = .18$, $p = .86$). This suggests that participants may be understanding the metaphor of marker color can relate to gender.

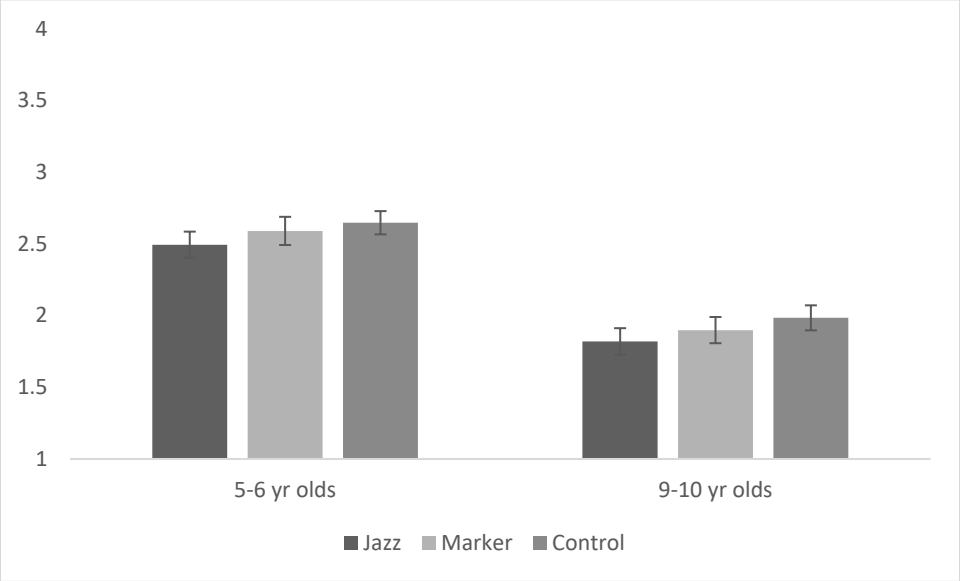
The GES-C by Condition and Age

Overall, there was no effect of hearing a story on the 16-item gender essentialism scale ($ps \geq .10$). Participants in the combined story conditions ($M = 2.20$, $SD = .60$) were not significantly different from participants in the control condition ($M = 2.35$, $SD = .56$; Stories vs. Control $t(169) = -1.55$, $p = .12$). There was no difference in essentialism between the control condition and either of the two story conditions (Jazz story $M = 2.17$, $SD = .59$, Marker story $M = 2.24$, $SD = .61$; Jazz vs. Control $t(168) = -1.65$, $p = .10$; Marker *vs.* Control $t(168) = -1.02$, $p = .31$). Nor was there a difference between the Jazz and Marker stories (Jazz vs. Marker $t(168) = -.63$, $p = .53$).

Neither younger nor older participants showed a difference in essentialism between the stories and the control (5-6-year-olds $ps \geq .25$; 9-10-year-olds $ps \geq .21$; see Figure 3.2). When I

looked at only participants who did not know a GNC person, there was still no difference in essentialism scores (*p*s ≥ .83).

Figure 3.2 Scores on the Full GES-C by Condition and Age Group

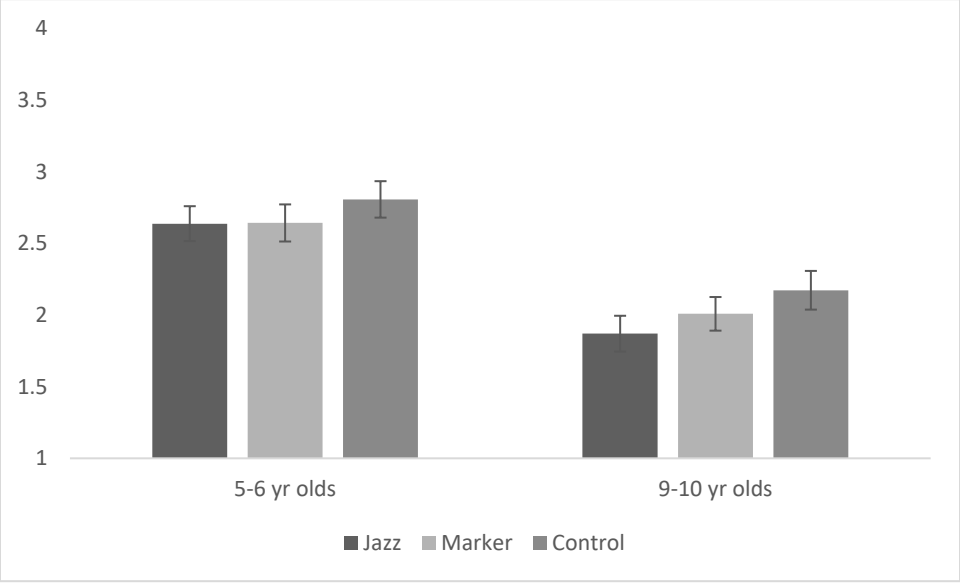


Note. Higher numbers signify more essentialism. Error bars represent standard errors for conditions and age category.

Examining the components, I found no differences in discreteness (*p*s ≥ .11) or informativeness (*p*s ≥ .12) by condition. When looking at age groups separately, there were also no differences in informativeness (5-6 -years-old *p*s ≥ .10; 9-10-year-olds *p*s ≥ .36). However, for younger children, those in the Jazz condition (*M* = 2.21, *SD* = .69) endorsed discreteness significantly less than those in the Marker condition (*M* = 2.66, *SD* = .67; Jazz vs. Marker *t*(86) = -2.33, *p* = .02). There were no other significant differences between conditions in endorsing discreteness for younger children (*p*s ≥ .23) and there were no differences between conditions for older children (*p*s ≥ .57).

Turning to the biological basis component, I found participants in the two stories collapsed into one condition (M = 2.29, SD = .74) were trending lower, but not significantly so, than the control condition (M = 2.52, SD = .77; Stories vs. Control $t(169) = -1.89$, $p = .06$). Those in the Jazz condition (M = 2.27, SD = .75) did not endorse biological basis beliefs significantly less than the control condition (Jazz vs. Control $t(112) = -1.82$, $p = .07$), and those in the Jazz condition and Marker condition (M = 2.32, SD = .73) did not differ from one another (Jazz vs. Marker $t(111) = -.37$, $p = .71$). Those in the Marker condition were not significantly different from those in the control condition (Marker vs. Control $t(113) = -1.45$, $p = .15$). When I looked at each age group, I found no differences between participants' scores by condition (5-6 -years-old ps $\geq .28$; 9-10-year-olds ps $\geq .10$); see Figure 3.3).

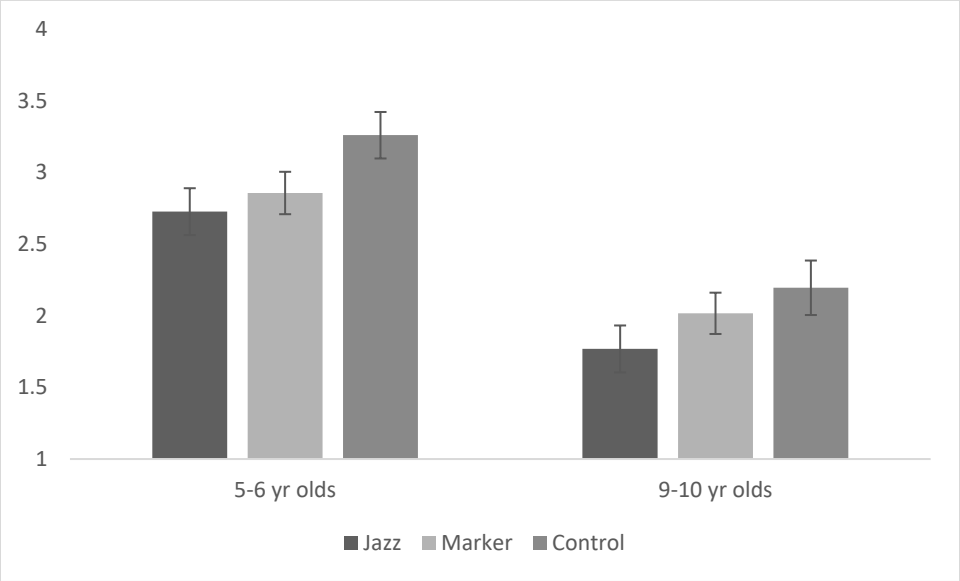
Figure 3.3 The GES-C Biological Basis Scores by Condition and Age Group



Note. Higher numbers signify more essentialism. Error bars represent standard errors for conditions and age category.

Where I did obtain significant differences was on the immutability component. Overall, participants in both story conditions combined (M = 2.35, SD = .93) endorsed less immutability than those in the control condition (M = 2.78, SD = 1.08; Stories vs. Control $t(101.91) = -2.62$, p = .01). Those in the Jazz condition (M = 2.26, SD = .99) expressed less immutability than those in the control condition (Jazz vs. Control $t(112) = -2.68$, p = .01) but did not significantly differ from those in the Marker condition (M = 2.43, SD = .88; Jazz vs. Marker $t(111) = -.94$, p = .35). Those in the Marker condition endorsed immutability non-significantly less than those in the control condition (Marker vs. Control $t(109.48) = -1.93$, p = .06). For younger children, participants in the two story conditions combined (M = 2.79, SD = .83) endorsed immutability beliefs less than those in the control condition (M = 3.26, SD = .92; Stories vs. Control $t(87) = -2.47$, p = .02). Those in the Jazz story (M = 2.73, SD = .88) expressed less immutability than those in the control condition (Jazz vs. Control $t(86) = -2.41$, p = .02). Those in the Marker story (M = 2.86, SD = .78) endorsed immutability non-significantly less than those in the control condition (Marker vs. Control $t(86) = -1.80$, p = .08). There was no difference between the two story conditions (Jazz vs. Marker $t(86) = -.57$, p = .57). Older children's immutability scores did not differ across the two stories (p = .15). However, the Jazz story (M = 1.77, SD = .85) non-significantly reduced immutability when compared to the control condition (M = 2.20, SD = .87; Jazz *vs.* Control $t(51) = -1.71$, p = .09). There was no significant difference between the Marker story (M = 2.02, SD = .78) and the Control, $t(53) = -.76$, p = .45; see Figure 3.4 for means and standard deviations by condition and age.

Figure 3.4 The GES-C Immutability Scores by Condition and Age Group



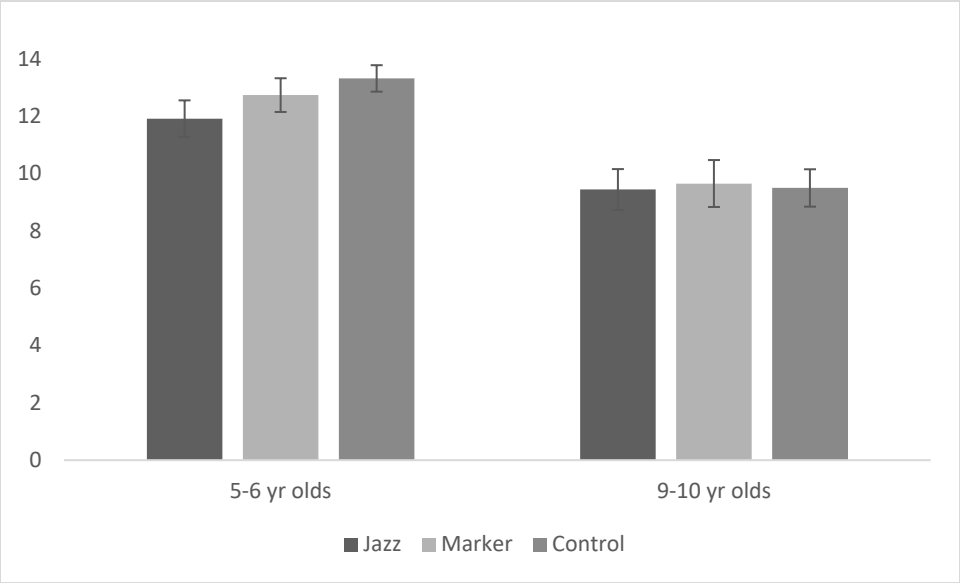
Note. Higher numbers signify more essentialism. Error bars represent standard errors for conditions and age category.

Island Task by Condition and Age

There were no differences in the Island Task results as a function of condition. Participants in the combined stories condition (M = 10.90, SD = 3.88) did not differ from participants in the control condition (M = 11.56, SD = 3.49) (Stories vs. Control $t(163) = -1.11$, $p = .27$). Participants in the Jazz story condition (M = 10.63, SD = 3.66) did not significantly from those in the Marker story condition (Jazz *vs.* Marker $t(106) = -.68$, $p = .50$) or control condition (M = 11.58, SD = 3.49; Jazz vs. Control $t(107) = -1.38$, $p = .17$). The Marker condition also did not differ significantly from the control condition (Marker vs. Control $t(111) = -.61$, $p = .54$). There were no differences between conditions for older children ($ps \geq .84$). There were no significant differences between conditions for the younger children, but those in the Jazz condition (M = 11.92, SD = 3.19) showed a non-significant tendency to be less essentialist than

those in the control condition (M = 13.32, SD = 2.57; Jazz vs. Control $t(54) = -1.82$, $p = .07$; see Figure 3.5). When I looked specifically at participants who did not know a GNC person, there were no differences between conditions (ps ≥ .71).

Figure 3.5 Scores on Island task by Condition and Age Group



Note. Higher numbers signify more essentialism and 15 is the highest score. Error bars represent standard errors for conditions by age category.

Discussion

This study suggests that a story regarding a transgender character, or a story that could be construed as a metaphor for a transgender character, can influence one component of children's gender essentialism, specifically immutability. However, for younger children especially, hearing a direct, realistic story was the only significant means of teaching children about transgender identities and reducing belief in gender immutability—and even here, effects were modest. This study revealed that stories can be a way to teach children about the social world and

change essentialist beliefs, but the impact may be limited and greatly affected by features of the story.

Although overall gender essentialism was not changed by the stories, immutability was reduced by exposure to a trans character. This may be a result of the story discussing the character's social transition. Change was a significant theme throughout both stories (e.g., "Then one amazing day, everything changed.") and was stressed in the questions asked by the experimenter. The biological basis component was also non-significantly trending towards being reduced by the stories. This may be because the stories stressed the difference between characters' bodies and the identities they felt to be on the inside, "I have a girl brain/red outside but a boy body/blue inside. I was born/made this way!" Past work has shown that children privilege "insides" when essentializing categories (Diesendruck & Weiss, 2015; Gelman & Wellman, 1991; Taylor et al., 2009), as a result we may have expected to find higher or unaffected rates of biological basis beliefs. On the other hand, showing a conflict between the inside and outside of the body may signal that biological features can vary independently of gender, and believing biological features can be independent of gender identity is an anti-essentialist belief. Given the non-significantly lower biological basis essentialism scores (p = .07), it remains an open question as to how children are interpreting the inside and outside explanation of gender identity outlined in the stories. That being said, the effects I found in my study seemed to be based in the most salient themes in the stories, however other themes were present in the stories as well.

Informativeness and discreteness were unaffected by the story. This may also be the result of the stories. In order to showcase the difference between each character's body and their feelings inside, the character's gender/color stereotypical behavior and preferences for their true

gender/color identity were emphasized (e.g., liking mermaids, drawing bluebirds). In fact, discussing the outside and inside differences is reminiscent of children's understanding that insides may be more important than outsides, for natural kinds (Gelman, 2004; Gelman & Wellman, 1991). So, although the stories were good at showcasing social transitions and the mutability of gender/color, this message may have been delivered by highlighting gender/color stereotypical behaviors and preferences that uphold existing essentialist beliefs.

If my theory about how the stories influenced and did not influence essentialism components is supported, a different story that does not highlight mutability at the expense of reinforcing beliefs about gender stereotypical behavior may be more successful at reducing gender essentialism more broadly. Since the publication of *I am Jazz* and *Red: A crayon's story*, many more children's books have been published that address gender diversity and other nonbinary and nonconforming identities. It is possible some of these stories may be more effective at reducing gender essentialism in children through indirect contact with different gender identities. Take for example, the book *When Aidan Became a Brother* by Kyle Lukoff. Aidan is a transgender boy. Before Aidan socially transitioned people thought he was a "different kind of girl" who did not like dresses. The author goes on to explain that many girls like stereotypically boy things. And Aidan is not shown as adhering strictly to boy stereotypes after his transition. The book then centers on Aidan's concern for making sure his soon-to-be-born new sibling will have the ability to explore and assert their gender, whichever identity that may be. He and his parents are careful not to gender the sibling. The nuances in Aidan's gender presentation, if perceived by young readers, may serve to reduce more than immutability beliefs but also beliefs about informativeness, discreteness, and biological basis. Exposing children to other gender identities like genderfluid or nonbinary identities may better illustrate

counterexamples to essentialist beliefs as well. The more an identity diverges from binary identities, the more beneficial contact with people with these identities may be for reducing gender essentialism in children. Developing stories that highlight identities that better challenge essentialist notions and introduce anti-essentialist beliefs related to more components may lead to more significant findings in essentialist reduction.

Another aspect of the study stories to consider is the role of acceptance from others. Not only did Jazz transition from a boy to a girl and Blue from a red marker to a blue marker, they both came to be supported by friends and family. The support modeled for participants in the story may have contributed to their acceptance of transgender identities. Furthermore, the friends in these stories could be ingroup members for participants, as participants were also school-aged children. This presents the opportunity for vicarious contact, a type of indirect contact that happens when another ingroup member has positive contact with an outgroup member. This contact has been explored through stories presented as videos and read aloud to school-aged children by experimenters; results of such studies showed a reduction in prejudice against outgroup members (Cocco et al., 2021; Husnu et al., 2018). A study conducted in Turkey with school-aged Turkish Cypriots found that a three-week intervention of hearing stories about friendships between Turkish Cypriot children and Greek Cypriot children resulted in Turkish children reporting more positive attitudes towards Greek Cypriot children, more intention to socialize with Greek Cypriot children, and more trust towards Greek Cyproit children (Husnu et al., 2018). This is notable due to a tendency toward animosity between Turkish Cypriots and Greek Cypriots. Similarly, in a study with Italian non-disabled school-aged children, children who saw a video of children playing with a child with a disability or were read a story about children playing with a child with a disability by an experimenter reported more positive

attitudes towards, more willingness to socialize with, and more willingness to help children with disabilities than children who did not see a video or hear a story (Cocco et al., 2021). Although my study does not explore prejudice, research shows a relation prejudice and essentialism (Ching et al., 2020; Ching & Xu, 2018; Fine et al., 2021; Haslam et al., 2002) and an effect of contact in reducing essentialist beliefs (Deeb et al., 2011; Fine et al., n.d.; Lytle et al., 2017b; Pauker et al., 2017). Therefore, the vicarious contact demonstrated in the Jazz and marker stories may have aided in reducing immutability beliefs for those participants in the experimental conditions.

Beyond the themes in the story, my study also contributes to the literature on how children extend information from fantastical and/or metaphorical stories versus more realistic, direct stories, when reasoning about the real world. Prior work showed that, even though it possible for children to learn from fantastical stories, it is more difficult because children may be overcautious with the information they transfer from these stories to real life (Geerdts et al., 2016; Richert & Smith, 2011; Walker & Lombrozo, 2017). Also, children's ability to understand metaphor develops over time (Di Paola et al., 2020; Gentner, 1988; Pearson, 1990) and certain discourse when using metaphors to teach science, for example, may be needed to help children understand metaphors (Cameron, 2002). One study found four- to six-year old-children had a more difficult time transferring information from a fantastical story to the real world than they did with a realistic story (Richert et al., 2009). I found that five- to six-year-old children in my study also were less likely to improve their understanding of transgender identities from the fantastical, metaphorical story. In contrast, nine- to ten-year-old children did seem to learn about transgender identities from both the fantastical, metaphorical and realistic stories. This indicates that it may be worth exploring when and how metaphors and fantastical story features can be effective in conveying messages about the social world.

This study also helps address a gap in the literature about what types of lessons children can learn from stories (Richert & Schlesinger, 2017). Much of the work about learning from stories focuses on more academic topics such as science and vocabulary. My work contributes to the small literature that focuses on social learning (Cocco et al., 2021; Husnu et al., 2018; Larsen et al., 2018; Mares et al., 2015; Strouse et al., 2018; Walker & Lombrozo, 2017). Given that so much of children's literature aims to teach children about the social world (e.g., *Ada Twist, Scientist* by Andrea Beaty and *Hair Love* by Matthew A. Cherry), it is important to study how children may learn from them.

Stories similar to the ones in this study also provide important indirect contact with transgender people. Earlier I noted that the transgender/GNC population is estimated to be very small (Meerwijk & Sevelius, 2017). Of the parents who were asked if their child had contact with someone who is GNC, only 21.3% reported their child knew someone who is GNC. Although this may seem like a large percentage given the estimate of the transgender population, we must also consider that each trans person knows many other non-trans people, making the percentage of the population that knows someone who is trans larger than the trans population itself. Another important consideration is that these parents volunteered their child to participate in a study that they were told was explicitly about how reading a story about a transgender character may influence children's beliefs about gender. Because this study focused on gender diversity, a topic that has been surrounded by debate and controversy in the media and in politics, this likely led to bias in who was interested in having their children participate.

Given the potential selection bias in my sample, it is also possible that the effects of the stories on gender essentialism found in this study may be specific to children having parents willing to have their child participate in discussions of gender and gender identities. Participants

in my study may have had conversations in the past around the topic of gender diversity. In fact, almost half of the parents who were asked, reported having heard of the books that inspired my study. As a result, these topics and types of storylines may not be completely new to some participants. This prior knowledge may result in participants being more affected by the messaging in the story because it is consistent with information they have received before and therefore reinforces their understanding of gender, and as a result, they continue to lower their essentialist beliefs. On the other hand, participants may be less affected by the messaging in the story because they have heard this information before and therefore have already adjusted their essentialist beliefs to account for this information leaving little room for more change. We may find different effects (either larger or smaller) with children who have more conservative parents or who come from more conservative areas, given the cultural differences in gender essentialism between children in liberal, urban areas and conservative, rural areas (Rhodes & Gelman, 2009a). These stories may also affect gender essentialism differently for children who have not been exposed to gender diversity. If children from conservative communities and families have not heard about transgender identities and, as past work has shown, express more gender essentialist beliefs, they may more easily reject some of their essentialist beliefs. However, given the environment in which children from conservative communities or families are being raised, they may not understand or acknowledge transgender identities from hearing a story and this may result in their gender essentialist beliefs being unchanged or even amplified as the story challenges the conceptualization of gender they have learned. This selection bias limitation may be difficult to avoid given the responsibility we have as researchers to inform participants about what our studies involve. However, there are interesting open questions about how effective

stories like those in this study are at changing essentialist beliefs in relation to children's prior exposure to gender diversity.

Although relatively few children in our study knew a GNC person, we were able to explore differences in essentialism between them and children in our study who did not know a GNC person. Those who knew a GNC person were significantly less gender essentialist and better understood what transgender identities were. This finding suggests a correlation between essentialism and contact, though the direction of causal influence is unclear. Although one possibility is that contact lowers essentialism, certainly it could be instead that lower essentialism leads to more contact with transgender people, or even that some third factor leads to both of these factors. Additionally, however, my study provides evidence for a causal relation of contact leading to less gender essentialism, given the experimental design. In any case, these correlations suggest it is worth examining if intergroup contact is effective in reducing gender essentialism, similar to how it is for adults (Fine et al., n.d.) and for reducing race essentialism (Deeb et al., 2011).

Another limitation of the study was the small sample size, with about 30 children per condition in each age group. Although this is an acceptable sample size in the developmental literature, a larger sample would be more statistically appropriate. We would expect to find at most modest effects, given that the stories were brief and children had 6 to 10 years of gender socialization. One solution to the issue of sample size would be to conduct additional similar studies and collect data from others' work similar to mine in order to conduct a meta-analysis.

Even with the obstacles to studying potentially sensitive subjects like gender identity, it is important to explore how we can teach young children about gender diversity. LGBTQ children face high rates of bullying based on their gender expression (J. Kosciw et al., 2018; J. G. Kosciw

et al., 2019). This struggle is obvious in the treatment of transgender people by politicians pushing for legislation that limit transgender people's access to places ranging from restrooms to the military (Fausset, 2017; Hersher & Johnson, 2017; Montgomery, 2017). My work contributes to research that supports inclusivity and examines how children can learn about important social concepts like gender diversity in the hopes that world can become a safe place for gender nonconforming and nonbinary people. This study suggests that introducing children to transgender characters through direct, realistic stories is one potential way to educate children and may reduce certain essentialist beliefs.

Chapter 4 General Discussion

In the two studies of this dissertation, I created and used the Gender Essentialism Scale for Children (GES-C) and examined how children's stories about transgender identities influence gender essentialism. I developed a theoretically based scale and tested it with a large sample of children, finding that it was reliable and valid. The GES-C was developed to measure four components of gender essentialism: biological basis, discreteness, informativeness, and immutability. Using the GES-C, I found that a realistic story that directly discusses a transgender child can reduce gender immutability beliefs in children.

These studies point to the continued need for research in the field of gender essentialism. This is especially true given the current political and societal discourse around gender and gender identities. Essentialist beliefs have long been linked to negative outcomes such as prejudice, stereotyping, and discrimination (Bastian & Haslam, 2006; Nick Haslam et al., 2006; Roberts et al., 2017; Wilton et al., 2019). A large body of research has also identified the trajectory of endorsement of gender essentialism beliefs for adults and children, as well as cultural factors that can alter this trajectory (Rhodes & Gelman, 2009a). What continues to be important is to harness this knowledge to reduce harmful essentialist beliefs and their downstream consequences.

The first step to addressing issues related to gender essentialism is to effectively and efficiently measure it. Prior to the GES-C, gender essentialism in children was measured in a piecemeal fashion. Researchers focused on particular components of gender essentialism in their research but not all components at once. This made it difficult to get a fuller picture of how

gender essentialism and its components operated. Furthermore, researchers did not always have an opportunity to test the reliability and validity of their essentialism measures beyond face value. Therefore, there are no prior measures of child essentialism with CFAs conducted using structural equation modeling, and many of these existing measures were presented without Cronbach's alpha statistics or validity analyses. The GES-C allows for all components to be analyzed separately and together in one short scale and analyses were conducted to investigate the scales' reliability and validity.

The GES-C was found to be valid and reliable in children. Cronbach's alpha scores indicated the scale was reliable. The components of the scale correlated with each other as expected. Furthermore, the GES-C and those components that are conceptually related to the Island Task correlated positively. Results from Chapter 3 revealed that children whose parent reported that their child had contact with a gender nonconforming person reported less gender essentialism than those who had not had that contact, and in Chapter 2 girls endorsed less gender essentialism than boys. Furthermore, gender essentialism as measured by the GES-C decreased with age in Chapters 2 and 3 These results are consistent with past research on contact and gender and sexual orientation essentialism (Fine et al., n.d.; Gülgöz et al., 2019; Lytle et al., 2017a), as well as past research on participant gender (Smiler & Gelman, 2008) and age (Taylor, 1996; Taylor et al., 2009) differences in gender essentialism.

Yet the results from the confirmatory factor analysis using structural equation modeling did not result in ideal fit statistics with a sample of children. The GES-C may benefit from a new reverse-coded item for the informativeness component. It is also possible that the measure was too complex for the sample size I was able to collect. Encouragingly, the fit statistics of a scale measuring children's ambivalent sexism were similar to those of the GES-C in my child sample

(Hammond & Cimpian, 2021), though both scales fell short of the more conservative cutoffs for acceptable fit in research with adults. It is an open question as to whether different cutoffs should be used with models with data from young children. My CFA results and those of Hammond and Cimpian (2021) demonstrate a need for researchers and statisticians to address CFAs conducted with child samples, due to the difficulty of recruiting samples and the developmental differences that may contribute to more error in measurement.

Reducing gender essentialism was the goal of Chapter 3, and I found that stories may modestly affect essentialism endorsement. This study examined how well elementary school-aged children learned about transgender identities from a direct, realistic story of a transgender girl who is going through a social transition, as well as a metaphorical, fantastical story about a marker who was red on the inside but blue on the outside and transitioned to an identity of a blue marker. However, the stories I used significantly reduced only immutability beliefs.

This may be due to the plot of the stories I chose. Being more intentional about addressing each component of gender essentialism may lead to different results. As I described in Chapter 3, the main characters highlighted ways in which they conformed (realistically or metaphorically) to stereotypical presentations of their gender. These stereotypical presentations may also highlight a discreteness between gender identities. Future work looking at how gender essentialism can be reduced through indirect contact may benefit from focusing on individual components or making sure all components are addressed within a study manipulation or intervention.

Although the mixture of essentialist and anti-essentialist messaging in the story may have led to an inconsistent reduction of essentialist beliefs in children, it does offer some evidence that the GES-C may be measuring different components. The fact that the four components of the

GES-C were not affected similarly suggests that the subscales may be capturing distinct components of gender essentialism. Specifically, the stories used in Chapter 3, which prominently featured identity transitions, may be most relevant to the component of immutability, and could explain why immutability was the only subscale of the GES-C that was significantly reduced by the stories. This is a promising finding for future work that may further help validate the subscales of the GES-C.

Beyond the messaging in the story, there are other elements to be cognizant of when considering new stories to reduce essentialism: metaphor and fantasy. Fantasy and metaphor are prevalent in children's stories but relatively little research has examined how children learn about the social world from such stories (Strouse et al., 2018). My results suggest that realistic stories may better convey messages about the social world. Yet fantastical stories are fun to read, and metaphors may also allow teachers and parents to address sensitive and controversial topics in a way that does not offend people. My study leaves many open questions for future studies. For example, is it necessary to have realistic protagonists but less consequential to have a fantastical setting when changing beliefs about the social world? What story features or adult scaffolding are necessary to help younger children understand metaphors? Answering these questions can contribute to creating effective, well-received interventions that avoid controversy.

Gender essentialism is a multi-faceted lay theory that can have negative downstream consequences related to prejudice and discrimination for both children and adults. The preceding two studies demonstrated the importance of assessing levels of essentialism comprehensively in order to gauge the effectiveness of interventions to reduce gender essentialism. My work revealed that the GES-C can comprehensively measure gender essentialism in children. With

future studies, the GES-C can be further improved and used in intervention work aimed at reducing gender essentialism.

Appendix A The GES-C Pilot Studies

I completed two rounds of piloting. In one round, I tested 42 preliminary items with 35 participants. I conducted an exploratory factor analysis (EFA) using structural equation modeling in MPlus (Muthén & Muthén, n.d.) and chose the highest loading items on each factor that also, taken at face value, corresponded with the theory of the components. I then conducted a second, larger pilot of 94 participants with the selected items and conducted another EFA. I made final changes and decisions to create the final items of the GES-C based on the EFA, the reliability, Cronbach's alphas, of the subscales, and face validity.

Method

Pilot 1

Participants

Participants were 35 children aged five to twelve years old (M = 7.73 yrs, SD = 1.81 yrs). Twenty-seven were girls (77.1%) and 8 (22.9%) were boys. No parent reported having a child with another gender. Sixteen (45.7%) of participants were White, 8 (22.9%) were Asian, 3 (8.6%) were multiracial, 2 (5.7%) were Black, and 1 (2.9%) was Middle Eastern, 5 (14.3%) did not have race data reported or the race information reported was illegible. Data was collected at local children's museums and in-lab at the University of Michigan.

Materials

For each question, participants are asked to respond on a 6-point scale from 'Disagree a lot' to 'Agree a lot' scale illustrated by circles getting increasingly larger, as continuous

measures have been found to be more reliable than dichotomous scales in adults (Cohen, 1983; Greenwald & O'Connell, 1970; Stöber et al., 2002). Higher numbers on the scale represent more essentialist beliefs. Below are the instructions to train participants to use the circle scale shown to children (see Table A.1 for items piloted).

Instructions

I am going to say some sentences and I want you to tell me how much you agree or disagree with them using this scale. You agree with something when you think it is true and you disagree with something when you DON'T think something is true.

See this circle? This means you Agree a lot. This circle means you DISagree a lot. And these circles are somewhere in the middle.

Can you point to the circle that means….? (GO THROUGH 2 SCALE POINTS)

[If child chooses wrong circle, correct them.]

Okay, let's practice a bit more. Remember to point to the circle that shows how much you agree or disagree with the sentences I say.

How much do you agree or disagree with this sentence? Broccoli is a good dessert.

How about, it's fun to go to play outside.

Remember that you can also answer somewhere in the middle. For example, if I were to say, flowers are yellow, I would answer somewhere in the middle (point to a middle circle) because some flowers are yellow and some aren't.

Let's try a couple more.

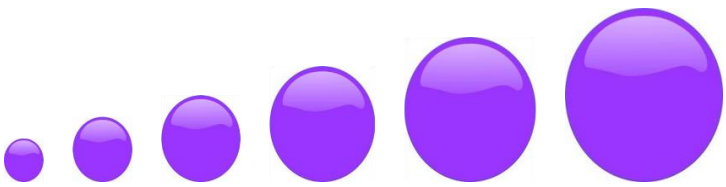
Cars are red.

Okay, one more, puppies grow up to be dogs.

Make a note if kid does not understand the scale or if they are being silly.

Looks like you've got it! Now let's move onto some more sentences...

To keep track of how much we've done, we'll use these stamps. Every time you finish a group of items, you can put an animal stamp in your zoo. I'll tell you when we finish a group.



Procedure

Participants were randomly assigned to answer one of four versions of the scale (see the Materials section in Chapter 2 for an explanation of the scale versions). Participants saw items in a random order. To help maintain participants' interest in the task, participants were given stamps and a zoo scene. Every seventh item they were able to stamp an animal in their zoo scene.

Results

I conducted an EFA and found that the four-factor solution resulted lowest chi-squared value. Although the model fit was poor, it was superior to the 1-, 2-, and 3-factor models generated by the EFA [RMSEA=.195(CI: .185, .205); CFI=.263; TLI=.092; $\chi^2(699) = 1977.81$, $p < .001$]. I selected four items from each of the four factors that mapped onto the four components of essentialism based on face value and loadings. I then calculated the Cronbach's alpha for the entire 16-item scale and the 4-item components (Full scale $\alpha = .71$; Biological basis $\alpha = .27$; Discreteness $\alpha = .69$; Informativeness $\alpha = .43$; Immutability $\alpha = .83$; see Table 1 for item loadings and item selection).

Table A-.1 The Factor (latent variable) Loadings from the Preliminary EFA

Item	Informativeness 1	Discreteness 2	Immutability 3	Biological Basis 4
1. You can know a kid is a girl or boy when they are in their mom's tummy	.13	.10	.16	.26
2. You know a kid is a girl or boy because of what their bodies look like	.43	.21	.06	.41
3. You can't know if a kid is a boy or girl just by their bodies (R)	.01	.10	.12	.07
4. A kid's body is what makes them a girl or a boy	.03	.64	.40	.14
5. Scientists can tell if a kid is a girl or a boy just by looking at their blood	.06	.41	.35	.29
6. You have to learn how to be a girl or a boy (R)	.47	.13	.03	.04
7. Girls do girl things because they have girl bodies	.71	.10	.22	.34
8. Boys do boy things because they have boy bodies	.74	.25	.31	.13
9. A kid can have a boy body and act like a girl (R)	.27	.05	.18	.18
10. A kid can have a girl body and act like a boy (R)	.23	.57	.14	.04

11. A kid can have a boy body and be a girl (R)	.62	.21	.65	.12
12. A kid can have a girl body and be a boy (R)	.13	.04	.73	.04
13. Being a boy or girl is like being short or tall, a kid can't decide what they will	.03	.44	.26	.01
14. A kid is a boy or girl because of how they think, not what their body looks like (R)	.19	.16	.44	.23
15. A kid is either a girl or a boy	.06	.18	.24	.06
16. Girls are always girls, even when they do boy things	.22	.17	.23	.04
17. Girls do mostly girl things	.64	.08	.27	.49
18. A boy can do girl things (R)	.47	.53	.07	.04
19. A kid can only act like a girl or a boy; they can't act like a girl and a boy	.48	.23	.02	.09
20. A kid can feel like a boy and a girl at the same time (R)	.39	.40	.24	.13
21. A kid can feel like a boy sometimes and girl other times (R)	.19	.79	.21	.10
22. Sometimes it is hard to tell if a kid is a boy or a girl (R)	.05	.33	.21	.78
23. It's easy to tell if a kid is a boy or a girl	.27	.05	.46	.26

24. A kid can be a girl and a boy (R)	.27	.23	.72	.09
25. Boys and girls are a lot alike (R)	.07	.01	.15	.35
26. Boys and girls are very different from one another	.52	.12	.29	.12
27. If a kid is a girl, you know they like girl things	.61	.39	.36	.09
28. You know a lot about a kid when you know they are a girl	.35	.48	.12	.06
29. Just because you know a kid is a girl doesn't mean you know a lot about them (R)	.35	.11	.17	.06
30. It's hard to tell if a kid is a boy or girl by how they act (R)	.10	.43	.06	.14
31. It's hard to tell if a kid is a boy or girl by the toys they like (R)	.17	.06	.03	.40
32. It's easy to tell if a kid is a boy or a girl by how they act	.29	.50	.01	.13
33. It's easy to tell if a kid is a boy or a girl by the toys they like	.58	.16	.09	.16
34. There are some things only boys do	.26	.04	.10	.18
35. Boys all over the world act the same	.45	.22	.13	.47
36. Girls act like girls even when they're babies	.44	.13	.54	.07
37. A kid can change from a boy to a girl (R)	.42	.32	.74	.32
38. A kid can't change from a boy to a girl	.08	.05	.36	.76

39. A girl can act like a boy, but they are still a girl	.05	.27	.38	.21
40. A kid is born a girl or boy and that can’t change	.55	.26	.51	.14
41. A boy can become a girl if he wants (R)	.26	.16	.91	.08
42. If a kid is a girl when they're born, they will be a girl when they grow up, too	.32	.09	.55	.48

Note. The items highlighted in each factor are the items chosen to pilot in the second pilot for each component.

Discussion

With a very limited sample, I was able to narrow down my items in order to continue piloting. Although the sample was not sufficient for a structural equation model (Wolf et al., 2013), I conducted this analysis to maximize the usefulness of the results compared with the effort and time needed to recruit young children. Also, I had a large set of items and it is difficult to maintain a young child's attention through all of the items. I continued piloting items with the 16 items I selected with a larger sample and included the Island Task to help gauge the validity of my scale.

Pilot 2

With the 16 items chosen from the first pilot data, I conducted a second pilot with more participants and the addition of the Island Task (Taylor et al., 2009) described in the introduction of this study. I also changed the scale to a 4-point scale from 1 "Disagree a lot" to 4 "Agree a lot" that uses a two-step procedure to capture participants' responses, which makes using the scale easier for the youngest participants.

Method

Participants

I recruited 94 participants. Participants were children aged five to twelve years old (M = 7.83 yrs, SD = 1.94 yrs). Fifty-five were girls (58.5%) and 39 (41.5%) were boys. No parent reported having a child with another gender. Thirty-nine (41.5%) of participants were White, 9 (9.6%) were multiracial, 3 (3.2%) were Latino/a, 2 (2.1%) were Asian, 2 (2.1%) were Black, and 1 (1.1%) was Middle Eastern, 38 (40.4%) did not have race data reported or the race information reported was illegible. Data was collected at local children's museums and in-lab at the University of Michigan.

Materials

Below are the instructions to train participants to use the thumb scale.

I am going to say some sentences and I want you to tell me how much you agree or disagree with them using this scale. You agree with something when you think it is true and you disagree with something when you DON'T think something is true.

See the thumbs up? This means you Agree. The thumbs down means you DISagree.

Can you point to the thumb that means….? (GO THROUGH 2 SCALE POINTS) If child chooses wrong thumb, correct them.

Okay, let's practice a bit more. Remember to point to the thumb that shows how much you agree or disagree with the sentences I say. How much do you agree or disagree with this sentence?

Broccoli is a good dessert.
Do you (dis)agree a little or do you (dis)agree a lot? (make hand gestures for a little or a lot with younger children)

How about, it's fun to go to play outside.

And, flowers are yellow.

(If answers (dis)agree a lot) Remember that you can also answer a little. For example, if I were to say, flowers are yellow, I would answer (dis)agree a little because some flowers are yellow, and some aren't.
(If answers (dis)agree a little) Right, because some flowers are yellow and some aren't.

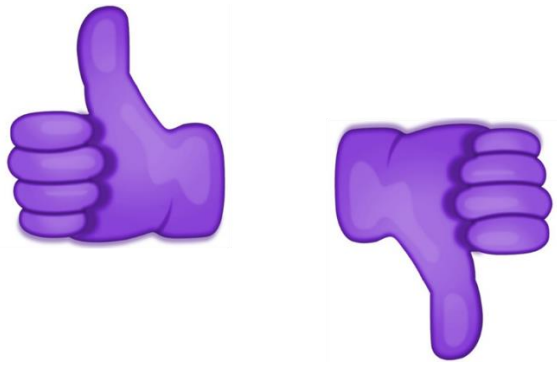
Let's try a couple more.

Cars are red. (Correct if needed) Remember that you can also answer a little. For example, if I were to say, cars are red, I would answer (dis)agree a little because some cars are red, and some aren't.

Okay, one more, puppies grow up to be dogs. (Make a note if kid does not understand the scale or if they are being silly.)

Looks like you've got it! Now let's move onto some more sentences...
To keep track of how much we've done, we'll use these stamps. Every time you finish a group of items, you can put an animal stamp in your zoo. I'll tell you when we finish a group.

How much do you agree or disagree with this sentence? Remember, thumbs up means agree, and thumbs down means disagree.



Measures

Pilot 2 Gender Essentialism Items. Participants agreed or disagreed with each of the 16 items selected from Pilot 1 and then indicated whether agreed or disagreed "a little" or "a lot." This resulted in a 4-point scale from "disagree a lot" to "agree a lot." Items were averaged together and scores ranged from 1 to 4 with higher numbers indicating more essentialism.

Island Task (Taylor et al., 2009)**.**

This task was described in Chapter 2. There are 15 items in total and participant answers that reflect the gender assigned the child (are stereotypical or expected based on the gender of the baby now 10 yr old) are scored as "1" and the answers that reflect the child's environment (are stereotypical or expected based on the gender of the other people on the island) are scored "0." Scores can range from 0 to 15 with higher numbers indicating more essentialism.

Procedure

Like in Pilot 1, participants were randomly assigned to answer one of four versions of the scale (girls and boy vs. boys and girls by girl questions vs. boy questions). Participants saw items in a random order. After, participants completed the Island Task. Participants were given stamps and a zoo scene again. Every few items they were able to stamp an animal in their zoo scene.

Results

First, I attempted a confirmatory factor analysis based on the items selected in Pilot 1 but there was no convergence in the model. As a result, I conducted an EFA. There was no convergence for the two, three, or four factor models. When I removed the biological component, I was able to successfully run an EFA and found the three-factor model to fit the data and theory best [RMSEA=.077(CI: .028, .118); CFI=.954; TLI=.908; $\chi^2(33) = 50.58$, $p < .03$; see Table 2 for factor loadings].

Table A-.2 The Factor (latent variable) Loadings from Pilot 2 EFA without the Biological Basis Component

Item	Informativeness	Immutability	Discreteness
	1	2	3
Info 1. Girls all over the world act the same	.57	.09	.17
Info 2. It's easy to tell if a kid is a girl or a boy by how they act	1.14	.30	.35
Info 3. There are some things only girls do	.18	.08	.06
Info 4. It's easy to tell if a kid is a girl or a boy by the toys they like	.71	.38	.41
Dis 1. A girl can do boy things	.29	.34	.61
Dis 2. A kid can feel like a girl sometimes and a boy other times	.25	.37	.82
Dis 3 A kid can feel like a girl and a boy at the same time	.15	.10	.56
Dis 4. Girls do mostly girl things	.17	.24	.34
Immut 1. A kid can change from a girl to a boy	.14	.97	.43
Immut 2. A girl can become a boy if she wants	.21	.80	.38
Immut 3. A kid is born a girl or boy and that can't change	.22	.58	.20

Immut 4. If a kid is a girl when they're born, they will be a girl when they grow up, too	.16	.55	.22

Note. Info = Informativeness, Dis = Discreteness, Immut = Immutability.

I also computed the reliability of the scale. These 16 items were reliable (α = .76), but the some of the components had lower reliability (Biological basis α = .44; Discreteness α = .63, Informativeness α = .56, Immutability α = .82) and not all components had reverse coded items. The biological basis component was particularly low, so I computed the reliability for the scale without the biological component and the reliability improved (α = .78).

The 16-item scale negatively correlate with age (r = -.42, p < .001). However, the scale did not correlate with the Island Task (r = .19, p = .14).

Discussion

The EFA and Cronbach's alphas revealed that the three components of discreteness, informativeness, and immutability were acceptable factors. The biological component, however, did not show reliability or acceptable coherence as a factor. As a result of the findings, I made several adjustments to the scale.

Biological Basis. I was advised that the biological basis items mentioning scientists and doctors may signal expertise and knowledge to children that may affect how children respond to these items (Fitneva, 2010; Landrum & Mills, 2015; Lutz & Keil, 2002; Toyama, 2017). This may have contributed to the lower reliability found with this component. Therefore, I replaced items that mentioned scientists and doctors.

Discreteness. The discreteness items about doing "boy/girl things" may be to conceptually similar to informativeness so it was replaced.

Informativeness. The informativeness component did not have a reverse-coded item so I replaced "There are some things only girls do" with a reverse-coded item.

Immutability. The immutability item "A kids is born a girl or boy and that can't change" may be conceptually similar to the biological basis component because the item suggests that change cannot occur for a biological reason, being born.

There are 4 versions of the scale for the gender used in single gender items and order of gender in two gender items:

1) Girls with girls and boys
2) Boys with girls and boys
3) Girls with boys and girls
4) Boys with boys and girls

Below are the final items selected to be used in Chapter 2:

Biological Basis

Bio1 A kid can choose to be a boy or a girl; it doesn't matter what their body looks like (reverse-coded; replaces Scientists can tell if a kid is a girl or a boy just by looking at their blood).

Bio2 Girls do girl things because they have girl bodies.

Bio3 You know a kid is a girl or boy because of what their body looks like.

Bio4 A girl is a girl, even before she is born (replaces Doctors can tell a kid is a girl or boy when they are in their mom's tummy).

Discreteness

Dis1 Girls and boys are the same (reverse-coded; replaces A girl can do boy things).

Dis2 A kid can feel like a girl sometimes and a boy other times (reverse-coded).

Dis3 A kid can feel like a girl and a boy at the same time (reverse-coded).

Dis4 Girls and boys are opposites (replaces Girls do mostly girl things).

Informativeness

Info1 Girls all over the world act the same.

Info2 It's easy to tell if a kid is a girl or a boy by how they act.

Info3 Girls all act very different from one another (reverse-coded; replaces There are some things only girls do).

Info4 It's easy to tell if a kid is a girl or a boy by the toys they like.

Immutability

Imut1 A kid can change from a girl to a boy (reverse-coded).

Imut2 A girl can become a boy if she wants (reverse-coded).

Imut3 A boy can't change that he is a boy.

Imut4 If a kid is a girl when they're born, they will be a girl when they grow up, too

Appendix B The 15-Item GES-C

Table B.1 Means and Standard Deviations for 15-item Full Scale and 3-item Informativeness Component by Study Sample

	Study1	Study 2	Study 3
15-item Full Scale	2.53(.54)	2.32(.59)	2.38(.60)
3-item Informativeness	1.50(.63)	1.81(.72)	1.87(.71)

Table B.2 Means and Standard Deviations for 15-item Full Scale and 3-item Informativeness Component by Scale Version

	Girl Boy Order	Boy Girl Order	Girl Questions	Boy Questions
15-item Full Scale	2.30(.60)	2.45(59)	2.34(.58)	2.41(.61)
3-item Informativeness	1.82(.73)	1.88(.69)	1.78(.65)	1.91(.77)

Bibliography

Aboud, F. E., & Brown, C. S. (2012). Positive and negative intergroup contact among children and its effect on attitudes. In *Advances in Intergroup Contact* (pp. 176–199). https://doi.org/10.4324/9780203095461

Ainsworth, C. (2015). Sex redefined. *Nature*, *518*(7539), 288–291. https://doi.org/10.1038/518288a

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley Pub. Co.

Arthur AE, Bigler RS, Liben LS, Gelman SA, & Ruble DN. (2008). Gender stereotyping and prejudice in young children. In *Intergroup attitudes and relations in childhood through adulthood* (pp. 66–86).

Bastian, B., & Haslam, N. (2006). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology*, *42*(2), 228–235. https://doi.org/10.1016/j.jesp.2005.03.003

Bem, S. L. (1983). Gender schema theory and its implications for child development: Raising gender-aschematic children in a gender-schematic society. *Signs: Journal of Women in Culture and Society*, *8*(4), 598. https://doi.org/10.1086/494131

Bem, S. L. (1989). Genital knowledge and gender constancy in preschool children. *Child Development*, *60*, 649–662.

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudices. *Advances in Child Development and Behavior*.

Booth, A. E. (2014). Conceptually coherent categories support label-based inductive generalization in preschoolers. *Journal of Experimental Child Psychology*, *123*(1), 1–14. https://doi.org/10.1016/j.jecp.2014.01.007

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research, Second Edition*.

Bruckmüller, S., Hegarty, P., & Abele, A. E. (2012). Framing gender differences: Linguistic normativity affects perceptions of power and gender stereotypes. *European Journal of Social Psychology*, *42*(2), 210–218. https://doi.org/10.1002/ejsp.858

Buck, D. M. (2016). Defining transgender: What do lay definitions say about prejudice? *Psychology of Sexual Orientation and Gender Diversity*, *3*(4), 465–472. https://doi.org/10.1037/sgd0000191

Cameron, L. (2002). Metaphors in the learning of science: A discourse focus. *British Educational Research Journal*, *28*(5), 673–688. https://doi.org/10.1080/0141192022000015534

Cao, C., & Meng, Q. (2020). Functional roles of mediated and extended mediated contact in promoting intergroup attitudes and post-conflict forgiveness: Mediating effects of cultural affinity. *Current Psychology, Allport 1954*. https://doi.org/10.1007/s12144-020-00612-8

Chestnut, E. K., Zhang, M. Y., & Markman, E. M. (2021). "Just as good": Learning gender stereotypes from attempts to counteract them. *Developmental Psychology, 57*(1), 114–125.

https://doi.org/10.1037/dev0001143
Ching, B. H. H., & Xu, J. T. (2018). The Effects of Gender Neuroessentialism on Transprejudice: An Experimental Study. *Sex Roles*, *78*(3–4), 228–241. https://doi.org/10.1007/s11199-017-0786-3
Ching, B. H. H., Xu, J. T., Chen, T. T., & Kong, K. H. C. (2020). Gender Essentialism, Authoritarianism, Social Dominance Orientation, and Filial Piety as Predictors for Transprejudice in Chinese People. *Sex Roles*, *2018*. https://doi.org/10.1007/s11199-020-01123-3
Chokshiaug, N. (2017, August 25). Canada Introduces 'X' as a Third Sex Category for Passport Holders. *New York Times*, A9.
Cocco, V. M., Bisagno, E., Di Bernardo, G. A., Cadamuro, A., Riboldi, S. D., Crapolicchio, E., Trifiletti, E., Stathi, S., & Vezzali, L. (2021). Comparing story reading and video watching as two distinct forms of vicarious contact: An experimental intervention among elementary school children. *British Journal of Social Psychology*, *60*(1), 74–94. https://doi.org/10.1111/bjso.12404
Cohen, J. (1983). The Cost of Dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253. https://doi.org/10.1177/014662168300700301
David, E. (2017). Capital T. *TSQ: Transgender Studies Quarterly*, *4*(1), 28–44. https://doi.org/10.1215/23289252-3711517
Davidson, N. S., & Gelman, S. A. (1990). Inductions from novel categories: The role of language and conceptual structure. *Cognitive Development*, *5*(2), 151–176. https://doi.org/10.1016/0885-2014(90)90024-N
Davoodi, T., Soley, G., Harris, P. L., & Blake, P. R. (2020). Essentialization of Social Categories Across Development in Two Cultures. *Child Development*, *91*(1), 289–306. https://doi.org/10.1111/cdev.13209
Deeb, I., Segall, G., Birnbaum, D., Ben-Eliyahu, A., & Diesendruck, G. (2011). Seeing isn't believing: The effect of intergroup exposure on children's essentialist beliefs about ethnic categories. *Journal of Personality and Social Psychology*, *101*(6), 1139–1156. https://doi.org/10.1037/a0026107
Di Paola, S., Domaneschi, F., & Pouscoulous, N. (2020). Metaphorical developing minds: The role of multiple factors in the development of metaphor comprehension. *Journal of Pragmatics*, *156*, 235–251. https://doi.org/10.1016/j.pragma.2019.08.008
Diesendruck, G., & Eldror, E. (2011). What children infer from social categories. *Cognitive Development*, *26*(2), 118–126. https://doi.org/10.1016/j.cogdev.2010.11.001
Diesendruck, G., & Gelman, S. A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin and Review*, *6*(2), 338–346. https://doi.org/10.3758/BF03212339
Diesendruck, G., Goldfein-Elbaz, R., Rhodes, M., Gelman, S., & Neumark, N. (2013). Cross-cultural differences in children's beliefs about the objectivity of social categories. *Child Development*, *84*(6), 1906–1917. https://doi.org/10.1111/cdev.12108
Diesendruck, G., & Haber, L. (2009). God's categories: The effect of religiosity on children's teleological and essentialist beliefs about categories. *Cognition*, *110*(1), 100–114. https://doi.org/10.1016/j.cognition.2008.11.001
Diesendruck, G., & HaLevi, H. (2006). The role of language, appearance, and culture in children's social category-based induction. *Child Development*, *77*(3), 539–553. https://doi.org/10.1111/j.1467-8624.2006.00889.x

Diesendruck, G., & Menahem, R. (2015). Essentialism promotes children's inter-ethnic bias. *Frontiers in Psychology*, *6*(August), 1–8. https://doi.org/10.3389/fpsyg.2015.01180

Diesendruck, G., & Weiss, E. (2015). Children's differential weighting of cues to social categories. *Cognitive Development*, *33*, 56–72. https://doi.org/10.1016/j.cogdev.2014.06.001

Elischberger, H. B., Glazier, J. J., Hill, E. D., & Verduzco-Baker, L. (2016). "Boys Don't Cry"—or Do They? Adult Attitudes Toward and Beliefs About Transgender Youth. *Sex Roles*, *75*(5–6), 197–214. https://doi.org/10.1007/s11199-016-0609-y

Emmerich, W., Goldman, K. S., & Kirsh, B. (1977). Evidence for a Transitional Evidence for a Transitional Phase in the Development of Gender Constancy. *Child Development*, *48*(3), 930–936.

Fausset, R. (2017). Deal on North Carolina Bathroom Law Would Expand Transgender Protections. *The New York Times*.

Fine, R. D., Gelman, S. A., Gulgoz, S., Horton, R., & Olson, K. R. (2021). Gender Essentialism in Rural and Urban Children Predicts Prejudice Against Gender Nonconformity. *Society for Research in Child Development 2021 Biennial Meeting*.

Fine, R. D., Gelman, S. A., & Ho, A. K. (n.d.). *Changing Beliefs about Gender: The Relation Between Contact with Gender Nonconforming Individuals and Reduced Gender Essentialism*.

Fitneva, S. A. (2010). Children's representation of child and adult knowledge. *Journal of Cognition and Development*, *11*(4), 458–484. https://doi.org/10.1080/15248371003700023

Flores, A., Gossett, C., Magni, G., & Reynolds, A. (2020). 11 openly LGBTQ lawmakers will take their seats in the next Congress . That ' s a record in both numbers and diversity . *Washington Post*. https://www.washingtonpost.com/politics/2020/11/30/11-lgbtq-legislators-will-take-their-seats-next-congress-largest-most-diverse-group-ever/

Ganea, P., Pickard, M. B., & DeLoache, J. (2008). Transfer between picture books and the real world by very young children. *Journal of Cognition and Development*, *9*(1), 46–66. https://doi.org/10.1080/15248370701836592

Geerdts, M. S., Van de Walle, G. A., & LoBue, V. (2016). Learning About Real Animals From Anthropomorphic Media. *Imagination, Cognition and Personality*, *36*(1), 5–26. https://doi.org/10.1177/0276236615611798

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* (Oxford Ser). Oxford University Press.

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, *8*(9), 404–409. https://doi.org/10.1016/j.tics.2004.07.001

Gelman, S. A., & Coley, J. D. (1990). The Importance of Knowing a Dodo Is a Bird: Categories and Inferences in 2-Year-Old Children. *Developmental Psychology*, *26*(5), 796–804. https://doi.org/10.1037/0012-1649.26.5.796

Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*, *57*(2), 396. https://doi.org/10.2307/1130595

Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology*, *66*(3), 327–353. https://doi.org/10.1016/j.cogpsych.2013.02.001

Gelman, S. A., Heyman, G. D., & Legare, C. H. (2007). Developmental changes in the coherence of essentialist beliefs about psychological characteristics. *Child Development*, *78*(3), 757–774. https://doi.org/10.1111/j.1467-8624.2007.01031.x

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*(3), 183–209. https://doi.org/10.1016/0010-0277(86)90034-X
Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, *38*(3), 213–244. https://doi.org/10.1016/0010-0277(91)90007-Q
Gentner, D. (1988). Metaphor as structure mapping : The relational shift. *Child Development*, *59*(1), 47–59.
Gouze, K. R., & Nadelman, L. (1980). Constancy of Gender Identity for Self and Others in Children between the Ages of Three and Seven. *Child Development*, *51*(1), 275. https://doi.org/10.2307/1129622
Graham, S. A., & Diesendruck, G. (2010). Fifteen-month-old infants attend to shape over other perceptual properties in an induction task. *Cognitive Development*, *25*(2), 111–123. https://doi.org/10.1016/j.cogdev.2009.06.002
Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, *37*(4), 827–838. https://doi.org/10.1177/001316447703700403
Greenwald, H. J., & O'Connell, S. M. (1970). Comparison of Dichotomous and Likert Formats. *Psychological Reports*, *27*(2), 481–482. https://doi.org/10.2466/pr0.1970.27.2.481
Guerra, R., Rebelo, M., Monteiro, M. B., Riek, B. M., Mania, E. W., Gaertner, S. L., & Dovidio, J. F. (2010). How should intergroup contact be structured to reduce bias among majority and minority group children? *Group Processes and Intergroup Relations*, *13*(4), 445–460. https://doi.org/10.1177/1368430209355651
Gülgöz, S., DeMeules, M., Gelman, S. A., & Olson, K. R. (2019). Gender essentialism in transgender and cisgender children. *PLoS ONE*, *14*(11), 1–12. https://doi.org/10.1371/journal.pone.0224321
Hall, M. (2015). *Red: A Crayon's Story*. Greenwillow Books.
Hammond, M. D., & Cimpian, A. (2021). "Wonderful but Weak": Children's Ambivalent Attitudes Toward Women. *Sex Roles*, *84*(1–2), 76–90. https://doi.org/10.1007/s11199-020-01150-0
Haslam, N, Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *The British Journal of Social Psychology*, *39 (Pt 1)*, 113–127. https://doi.org/10.1348/014466600164363
Haslam, Nick, Bastian, B., Bain, P., & Kashima, Y. (2006). Psychological essentialism, implicit theories, and intergroup relations. *Group Processes and Intergroup Relations*, *9*(1), 63–76. https://doi.org/10.1177/1368430206059861
Haslam, Nick, Rothschild, L., & Ernst, D. (2002). Are essentialist beliefs associated with prejudice? *British Journal of Social Psychology*, *41*(1), 87–100. https://doi.org/10.1348/014466602165072
Heiphetz, L., Gelman, S. A., & Young, L. L. (2017). The perceived stability and biological basis of religious beliefs, factual beliefs, and opinions. *Journal of Experimental Child Psychology*, *156*, 82–98. https://doi.org/10.1016/j.jecp.2016.11.015
Hersher, R., & Johnson, C. (2017). Trump Administration Rescinds Obama Rule on Transgender Students' Bathroom Use. *NPR*. https://doi.org/http://dx.doi.org/10.1108/17506200710779521
Herthel, J., & Jennings, J. (2014). *I am Jazz*. Dial Book.
Hoffarth, M. R., & Hodson, G. (2018). When intergroup contact is uncommon and bias is strong: the case of anti-transgender bias. *Psychology and Sexuality*, *9*(3), 237–250.

https://doi.org/10.1080/19419899.2018.1470107
Horak, L. (2014). Trans on YouTube: Intimacy, Visibility, Temporality. *TSQ: Transgender Studies Quarterly, 1*(4), 572–585. https://doi.org/10.1215/23289252-2815255
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118
Husnu, S., Mertan, B., & Cicek, O. (2018). Reducing Turkish Cypriot children's prejudice toward Greek Cypriots: Vicarious and extended intergroup contact through storytelling. *Group Processes and Intergroup Relations, 21*(1), 178–192. https://doi.org/10.1177/1368430216656469
Hussak, L. J., & Cimpian, A. (2019). "It feels like it's in your body": How children in the United States think about nationality. *Journal of Experimental Psychology: General, 148*(7), 1153–1168. https://doi.org/10.1037/xge0000567
Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2018). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist, 74*(2), 171–193. https://doi.org/10.1037/amp0000307
Jaswal, V. K., Lima, O. K., & Small, J. E. (2009). Compliance, conversion, and category induction. *Journal of Experimental Child Psychology, 102*(2), 182–195. https://doi.org/10.1016/j.jecp.2008.04.006
Keil, F. C. (1989a). Spiders in the web of belief: The tangled relations between concepts and theories. *Mind & Language, 3*(1–2), 43–50.
Keil, F. C. (1989b). Transformations on Natural Kinds and Artifacts. In *Concepts, Kinds, and Cognitive Development* (pp. 183–194). The MIT Press.
Kelemen, D., Emmons, N. A., Seston Schillaci, R., & Ganea, P. A. (2014). Young Children Can Be Taught Basic Natural Selection Using a Picture-Storybook Intervention. *Psychological Science*, *25*(4), 893–902. https://doi.org/10.1177/0956797613516009
Kinzler, K. D., & Dautel, J. B. (2012). Children's essentialist reasoning about language and race. *Developmental Science*, *15*(1), 131–138. https://doi.org/10.1111/j.1467-7687.2011.01101.x
Kosciw, J. G., Clark, C. M., Truong, N. L., & Zongrone, A. D. (2019). *The 2019 National School Climate Survey*. https://www.glsen.org/sites/default/files/2020-11/NSCS19-111820.pdf
Kosciw, J., Greytak, E., Palmer, N., & Boesen, M. (2018). *The 2017 National School Climate Survey*. GLSEN. https://files.eric.ed.gov/fulltext/ED590243.pdf
Landrum, A. R., & Mills, C. M. (2015). Developing expectations regarding the boundaries of expertise. *Cognition*, *134*, 215–231. https://doi.org/10.1016/j.cognition.2014.10.013
Larsen, N. E., Lee, K., & Ganea, P. A. (2018). Do storybooks with anthropomorphized animal characters promote prosocial behaviors in young children? *Developmental Science*, *21*(3). https://doi.org/10.1111/desc.12590
Lee, K. Y., Reis, H. T., & Rogge, R. D. (2020). Seeing the World in Pink and Blue: Developing and Exploring a New Measure of Essentialistic Thinking about Gender. *Sex Roles*. https://doi.org/10.1007/s11199-020-01141-1
Levy, G. D., & Carter, D. B. (1989). Gender schema, gender constancy, and gender-role knowledge: The roles of cognitive factors in preschoolers' gender-role stereotype attributions. *Developmental Psychology, 25*(3), 444–449. https://doi.org/10.1037/0012-1649.25.3.444
Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology, 74*(6), 1421–

1436. https://doi.org/10.1037/0022-3514.74.6.1421
Lissitsa, S., & Kushnirovich, N. (2020). Coevolution between Parasocial Interaction in Digital Media and Social Contact with LGBT People. *Journal of Homosexuality*, *00*(00), 1–24. https://doi.org/10.1080/00918369.2020.1809891
Lutz, D. J., & Keil, F. C. (2002). Early Understanding of the Division of Cognitive Labor Author (s): Donna J . Lutz and Frank C . Keil Published by : Wiley on behalf of the Society for Research in Child Development Stable URL : http://www.jstor.org/stable/3696271 REFERENCES Linked refer. *Child Development*, *73*(4), 1073–1084.
Lytle, A., Dyar, C., Levy, S. R., & London, B. (2017a). Contact with bisexual individuals reduces binegativity among heterosexuals and lesbian women and gay men. *European Journal of Social Psychology*, *47*(5), 580–599. https://doi.org/10.1002/ejsp.2241
Lytle, A., Dyar, C., Levy, S. R., & London, B. (2017b). Essentialist beliefs: Understanding contact with and attitudes towards lesbian and gay individuals. *British Journal of Social Psychology*, *56*(1), 64–88. https://doi.org/10.1111/bjso.12154
Mares, M. L., Sivakumar, G., & Stephenson, L. (2015). From Meta to Micro: Examining the Effectiveness of Educational TV. *American Behavioral Scientist*, *59*(14), 1822–1846. https://doi.org/10.1177/0002764215596555
Martin, C. L. (1989). Children's Use of Gender-Related Information in Making Social Judgments. *Developmental Psychology*, *25*(No. 1), 80–88. https://doi.org/10.1037/0012-1649.25.1.80
Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, *128*(6), 903–933. https://doi.org/10.1037/0033-2909.128.6.903
Medin, D. L. (1989). Concepts and Conceptual Structure. *American Psychologist*, *44*(12), 1469–1481. https://doi.org/10.1037/0003-066X.44.12.1469
Medin, D., & Ortony, A. (1989). Psychological essentialism. In *Similarity and Analogical Reasoning* (pp. 179–195). https://doi.org/10.1017/cbo9780511529863.009
Meerwijk, E. L., & Sevelius, J. M. (2017). Transgender population size in the United States: A meta-regression of population-based probability samples. *American Journal of Public Health*, *107*(2), e1–e8. https://doi.org/10.2105/AJPH.2016.303578
Miller, L. R., & Grollman, E. A. (2015). The Social Costs of Gender Nonconformity for Transgender Adults: Implications for Discrimination and Health. *Sociological Forum*, *30*(3), 809–831. https://doi.org/10.1111/socf.12193
Montgomery, D. (2017). Texas Transgender Bathroom Bill Falters Amid Mounting Opposition. *The New York Times*. https://www.nytimes.com/2017/08/08/us/time-is-running-out-on-texas-bathroom-bill.html
Muthén, B., & Muthén, L. (n.d.). *Mplus: A General Latent Variable Modeling Program*.
Pauker, K., Ambady, N., & Apfelbaum, E. P. (2010). *Race and salience and essentaialist thinking in racial stereotype development*. Child Development. http://search.ebscohost.com/login.aspx?direct=true&db=tfh&AN=55203532&site=ehost-live#.TzXCWmlfda0.mendeley
Pauker, K., Carpinella, C., Meyers, C., Young, D. M., & Sanchez, D. T. (2017). The Role of Diversity Exposure in Whites' Reduction in Race Essentialism Over Time. *Social Psychological and Personality Science*, 194855061773149. https://doi.org/10.1177/1948550617731496
Pauker, K., Xu, Y., Williams, A., & Biddle, A. M. (2016). Race Essentialism and Social

Contextual Differences in Children's Racial Stereotyping. *Child Development*, *87*(5), 1409–1422. https://doi.org/10.1111/cdev.12592
Payne, E., & Smith, M. (2014). The Big Freak Out: Educator Fear in Response to the Presence of Transgender Elementary School Students. *Journal of Homosexuality*, *61*(3), 399–418. https://doi.org/10.1080/00918369.2013.842430
Pearson, B. Z. (1990). The Comprehension of Metaphor By Preschool Children. *Journal of Child Language*, *17*(1), 185–203. https://doi.org/10.1017/S0305000900013179
Pettigrew, T. F., Christ, O., Wagner, U., & Stellmacher, J. (2007). Direct and indirect intergroup contact effects on prejudice: A normative interpretation. *International Journal of Intercultural Relations*, *31*(4), 411–425. https://doi.org/10.1016/j.ijintrel.2006.11.003
Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751
Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, *109*(34), 13526–13531. https://doi.org/10.1073/pnas.1208951109
Rhodes, Marjorie, & Gelman, S. A. (2009a). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, *59*(3), 244–274. https://doi.org/10.1016/j.cogpsych.2009.05.001
Rhodes, Marjorie, & Gelman, S. A. (2009b). Five-year-olds' beliefs about the discreteness of category boundaries for animals and artifacts. *Psychonomic Bulletin and Review*, *16*(5), 920–924. https://doi.org/10.3758/PBR.16.5.920
Rhodes, Marjorie, Gelman, S. A., & Karuza, J. C. (2014). Preschool Ontology: The Role of Beliefs About Category Boundaries in Early Categorization. *Journal of Cognition and Development*, *15*(1), 78–93. https://doi.org/10.1080/15248372.2012.713875
Rhodes, Marjorie, & Mandalaywala, T. M. (2017). The development and developmental consequences of social essentialism. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(August). https://doi.org/10.1002/wcs.1437
Richert, R. A., & Schlesinger, M. A. (2017). The role of fantasy–reality distinctions in preschoolers' learning from educational video. *Infant and Child Development*, *26*(4), 1–17. https://doi.org/10.1002/icd.2009
Richert, R. A., Shawber, A. B., Hoffman, R. E., & Taylor, M. (2009). Learning From Fantasy and Real Characters in Preschool and Kindergarten. *Journal of Cognition and Development*, *10*(1–2), 41–66. https://doi.org/10.1080/15248370902966594
Richert, R. A., & Smith, E. I. (2011). Preschoolers' Quarantining of Fantasy Stories. *Child Development*, *82*(4), 1106–1119. https://doi.org/10.1111/j.1467-8624.2011.01603.x
Roberts, S. O., & Gelman, S. A. (2016). Can White Children Grow Up to Be Black ? Children ' s Can White Children Grow Up to Be Black ? Children ' s Reasoning About the Stability of Emotion and Race. *Developmental Psychology*, *52*(6), 887. https://doi.org/10.1037/dev000132
Roberts, S. O., Gelman, S. A., & Ho, A. K. (2017). So It Is, So It Shall Be: Group Regularities License Children's Prescriptive Judgments. *Cognitive Science*, *41*, 576–600. https://doi.org/10.1111/cogs.12443
Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. Cambridge University Press.
Ruble, D. N., Balaban, T., & Cooper, J. (1981). Gender constancy and the effects of sex-typed

televised toy commercials. *Child Development*, *52*(2), 667–673. https://doi.org/10.1111/j.1467-8624.1981.tb03094.x
Ruble, Diane N., Taylor, L. J., Cyphers, L., Greulich, F. K., Lurye, L. E., & Shrout, P. E. (2007). The role of gender constancy in early gender development. *Child Development*, *78*(4), 1121–1136. https://doi.org/10.1111/j.1467-8624.2007.01056.x
Shutts, K., Pemberton, C. K., Spelke, E. S., Roben, C. K. P., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development*, *14*(1), 35–62. https://doi.org/10.1080/15248372.2011.638686.Children
Sivo, S. A., Xitao, F. A. N., Witta, E. L., & Willse, J. T. (2006). The search for "optimal" cutoff properties: Fit index criteria in structural equation modeling. *Journal of Experimental Education*, *74*(3), 267–288. https://doi.org/10.3200/JEXE.74.3.267-288
Slaby, R. G., & Frey, K. S. (1975). Development of Gender Constancy and Selective Attention to Same-Sex. *Child Development*, *46*(4), 849–856.
Sliter, K. A., & Zickar, M. J. (2014). An IRT Examination of the Psychometric Functioning of Negatively Worded Personality Items. *Educational and Psychological Measurement*, *74*(2), 214–226. https://doi.org/10.1177/0013164413504584
Smiler, A. P., & Gelman, S. A. (2008). Determinants of gender essentialism in college students. *Sex Roles*, *58*(11–12), 864–874. https://doi.org/10.1007/s11199-008-9402-x
Smyth, K., Feeney, A., Eidson, R. C., & Coley, J. D. (2017). Development of essentialist thinking about religion categories in Northern Ireland (and the United States). *Developmental Psychology*, *53*(3), 475–496. https://doi.org/10.1037/dev0000253
Sobel, D. M., Bloom, P., Weisberg, D. S., & Goodstein, J. (2013). Young children are reality-prone when thinking about stories. *Journal of Cognition and Culture*, *13*(3–4), 383–407. https://doi.org/10.1163/15685373-12342100
Steinmetz, K. (2014, June). The transgender tipping point. *Time*. http://time.com/135480/transgender-tippingpoint/
Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing continuous and dichotomous scoring of the balanced inventory of desirable responding. *Journal of Personality Assessment*, *78*(2), 370–389. https://doi.org/10.1207/S15327752JPA7802_10
Strouse, G. A., Nyhout, A., & Ganea, P. A. (2018). The role of book features in young children's transfer of information from picture books to real-world contexts. *Frontiers in Psychology*, *9*(FEB), 1–14. https://doi.org/10.3389/fpsyg.2018.00050
Szkrybalo, J., & Ruble, D. N. (1999). "God made me a girl": Gender constancy judgements and explanations revisited. *Developmental Psychology*, *35*(2), 392–402.
Tarlowski, A. (2018). Ontological constraints in children's inductive inferences: Evidence from a comparison of inferences within animals and vehicles. *Frontiers in Psychology*, *9*(APR). https://doi.org/10.3389/fpsyg.2018.00520
Taylor, M. G. (1996). The Development of Children' s Beliefs about Social and Biological Aspects of Gender Differences. *Child Development*, *67*(4), 1555–1571.
Taylor, M. G., Rhodes, M., & Gelman, S. A. (2009). Boys Will Be Boys ; Cows Will Be Cows : Children ' s Essentialist Reasoning about Gender Categories and Animal Species. *Child Development*, *80*(2), 461–481.
Toyama, N. (2017). Development of the selection of trusted informants in the domain of illness. *Infant and Child Development*, *26*(6), 1–18. https://doi.org/10.1002/icd.2039
Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005).

Rigidity and flexibility of gender stereotypes in childhood: Developmental or differential? *Infant and Child Development*, *14*(4), 365–381. https://doi.org/10.1002/icd.399
Tropp, L. R., Mazziotta, A., & Wright, S. C. (2016). Recent Developments in Intergroup Contact Research: Affective Process, Group Status, and Contact Valence. In F. Barlow & C. Sibley (Eds.), *Cambridge Handbook of the Psychology of Prejudice* (pp. 463–480).
van Anders, S. M. (2015). Beyond Sexual Orientation: Integrating Gender/Sex and Diverse Sexualities via Sexual Configurations Theory. In *Archives of Sexual Behavior* (Vol. 44, Issue 5). Springer US. https://doi.org/10.1007/s10508-015-0490-8
Vezzali, L., Hewstone, M., Capozza, D., Trifiletti, E., & Bernardo, G. A. Di. (2017). Improving Intergroup Relations with Extended Contact among Young Children: Mediation by Intergroup Empathy and Moderation by Direct Intergroup Contact. *Journal of Community and Applied Social Psychology*, *27*(1), 35–49. https://doi.org/10.1002/casp.2292
Walker, C. M., Gopnik, A., & Ganea, P. A. (2015). Learning to Learn From Stories: Children's Developing Sensitivity to the Causal Structure of Fictional Worlds. *Child Development*, *86*(1), 310–318. https://doi.org/10.1111/cdev.12287
Walker, C. M., & Lombrozo, T. (2017). Explaining the moral of the story. *Cognition*, *167*, 266–281. https://doi.org/10.1016/j.cognition.2016.11.007
Weems, G. H., & Onwuegbuzie, A. J. (2001). The impact of midpoint responses and reverse coding on survey data. In *Measurement and Evaluation in Counseling and Development* (Vol. 34, Issue 3, pp. 166–176). https://doi.org/10.1080/07481756.2002.12069033
Wilton, L. S., Bell, A. N., Carpinella, C. M., Young, D. M., Meyers, C., & Clapham, R. (2019). Lay Theories of Gender Influence Support for Women and Transgender People's Legal Rights. *Social Psychological and Personality Science*, *10*(7), 883–894. https://doi.org/10.1177/1948550618803608
Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, *73*(6), 913–934. https://doi.org/10.1177/0013164413495237
Wölfer, R., Christ, O., Schmid, K., Tausch, N., Buchallik, F. M., Vertovec, S., & Hewstone, M. (2019). Indirect Contact Predicts Direct Contact: Longitudinal Evidence and the Mediating Role of Intergroup Anxiety. *Journal of Personality and Social Psychology*, *116*(2), 277–295. https://doi.org/10.1037/pspi0000146.supp
Woolley, J. D. (1997). Thinking about Fantasy : Are Children Fundamentally Different Thinkers and Believers from Adults ? Published by : Wiley on behalf of the Society for Research in Child Development Stable URL : http://www.jstor.org/stable/1132282 Thinking about Fantasy : Ar. *Child Development*, *68*(6), 991–1011.
Woolley, J. D., & Ghossainy, M. E. (2013). Revisiting the fantasy-reality distinction: Children as naïve skeptics. *Child Development*, *84*(5), 1496–1510. https://doi.org/10.1111/cdev.12081
Worthington, R. L., & Whittaker, T. A. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. *The Counseling Psychologist*, *34*(6), 806–838. https://doi.org/10.1177/0011000006288127